

( 続紙 1 )

京都大学	博士 (情報学)	氏名	Yin Jou Huang
論文題目	Event Centric Approaches in Natural Language Processing (自然言語処理におけるイベント中心のアプローチ)		
(論文内容の要旨)			
<p>An overwhelmingly large amount of texts are published every day. These natural language texts encode the various events that happen around us, presenting a rich repository of events from which we can mine invaluable knowledge about the rapidly changing world. However, the large amount of data has gone far beyond human ability to read and process manually. Thus, how to automatically extract and analyze the events in natural language texts has become an important task. In this thesis, we study the events in natural language texts. Specifically, we focus on three representative event-related tasks: event coreference resolution, narrative event relation knowledge acquisition, and automatic summarization.</p> <p>In Chapter 1, we introduce the basics of events and the event-related tasks in natural language processing. Some of the tasks analyze individual events and aim to identify their composing factors, such as event extraction and event coreference resolution. Other tasks focus on the interactions and the various relations between multiple events, such as script learning and discourse parsing. In addition, many downstream tasks can benefit from incorporating events into the task, such as automatic summarization and question answering.</p> <p>In Chapter 2, we investigate the task of event coreference resolution, which is an important information extraction task. Event coreference resolution aims to identify and cluster the event mentions that refer to the same real-world event. In this work, we adopt the semantic view of events by representing events by a trigger and zero or more arguments of different semantic roles. For event mentions to be coreferent, their triggers must be semantically related, and their arguments must be compatible. While argument compatibility is an important aspect to determine the coreference status of event mentions, it is difficult to learn argument compatibility due to the limited size of existing annotated corpora. To tackle the above-mentioned problem, we propose a transfer learning framework of event coreference resolution and learn argument compatibility from the abundant unannotated corpora available. We conducted experiments on the benchmark KBP corpora and confirmed the effectiveness of our proposed method to capture argument compatibility.</p> <p>In Chapter 3, we study the task of narrative event relation knowledge acquisition. Narrative event relation captures the stereotypical ordering of events that are temporally and causally related, which is an important type of commonsense knowledge that reflects the way the human mind perceives the world. We adopt a 2-stage framework to extract narrative event relation knowledge from a large Japanese corpus. In the first stage, we utilize an association rule mining algorithm to identify narrative-related events of statistical significance. In the</p>			

second stage, we manually constructed an annotated corpus to learn the patterns of shared arguments between the event pairs. We propose a joint model which predicts the shared arguments and selectional preference simultaneously. With the methods above, we collected a large amount of narrative event relation knowledge from the web corpus.

In Chapter 4, we focus on the task of automatic summarization. Extractive summarization is one of the main paradigms of the automatic summarization task, which aims to identify the salient parts of a document for a summary. In this work, we consider the elementary discourse units (EDUs) as events and formulate extractive summarization as an event saliency identification problem. We consider two textual relations that are helpful for deciding the saliency of events: discourse relation and coreference relation. We propose a heterogeneous document graph to model the input document and the textual relations in it. Also, we propose a graph attention (GAT) based graph encoder to capture the structure of the heterogeneous document graph. We conducted experiments on the benchmark CNN/Daily Mail corpus and confirmed the effectiveness of our proposed method.

In Chapter 5, we conclude the event-centric approaches introduced in this thesis and discuss the directions of future research.

(続紙 2)

(論文審査の結果の要旨)

本論文は、自然言語テキストにおける意味表現・知識表現の基本単位であるイベントについて、イベントの構造解析、イベント間の関係解析、イベントに基づくアプリケーションなど、自然言語処理におけるイベント中心アプローチについて研究を行い、その成果をまとめたものである。得られた主要な成果は以下の通りである。

1. イベントの共参照解析について研究を行った。イベントは述語と項によって表現され、項の間の互換性がイベント間の共参照関係を定める重要な情報である。しかし、大規模なイベント共参照の訓練コーパスがないことから、その学習は困難であった。本研究では、生コーパスからイベント共参照の疑似正例と疑似負例を抽出することにより、項の間の互換性を学習する転移学習の手法を提案した。イベント共参照解析の標準データセットであるKBP2017に対する実験により、MUC/B3/CEAF<sub>Fe</sub>/BLANCの4つの評価指標の平均F1スコアにおいて既存手法を1.4ポイント上回る精度向上を実現した。

日本語コーパスから事態間関係知識の大規模抽出を行った。事態間関係知識は、よく共起するイベントのペアであり、その候補は相関ルールマイニングにより抽出することができる。しかし、日本語では項の省略が多いため、イベント間の共有項の識別を行うことが必要となる。本研究では、事態間関係知識のゴールドデータを構築し、これをもとに、項の対応付けと、述語の意味分類である格フレームの選択を同時学習する手法を考案した。構築したゴールドデータセットに対する実験で、既存手法に対して3.5ポイントのF1スコア上昇を確認した。

2. 抽出型文書要約に関する研究を行った。本研究では、抽出型文書要約をイベントの重要性認識タスクとして定式化し、文書を談話基本単位(Elementary Discourse Unit: EDU)に分割し、文書から重要性の高いEDUを要約として抽出する。文書中の談話関係と共参照関係を異種グラフとして表現し、グラフニューラルネットワークに基づく要約手法を提案した。文書要約の標準データセットであるCNN/Daily Mailにおいて評価指標ROUGE-1/Lで世界最高性能を実現し、提案手法の有効性を確認した。

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。令和3年6月17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、合格と認められた。また、本論文のインターネットでの全文公表についても支障がないことを確認した。