

様式VI

## 博士学位論文調査報告書

論文題目 Anchoring Events to the Time Axis toward Storyline Construction  
(ストーリーライン生成のための時間と事象情報の対応付け)

申請者氏名 坂口 智洋

最終学歴 平成 27年 3月  
京都大学大学院情報学研究科知能情報学専攻修士課程 修了  
平成 30年 3月  
京都大学大学院情報学研究科知能情報学専攻博士後期課程  
研究指導認定

学識確認 平成 年 月 日 (論文博士のみ)

論文調査委員 京都大学大学院情報学研究科  
(調査主査) 黒橋 禎夫 教授

論文調査委員 京都大学大学院情報学研究科  
西田 豊明 教授

論文調査委員 京都大学大学院教育学研究科  
楠見 孝 教授

( 続紙 1 )

京都大学	博士 (情報学)	氏名	坂口 智洋
論文題目	Anchoring Events to the Time Axis toward Storyline Construction (ストーリーライン生成のための時間と事象情報の対応付け)		
(論文内容の要旨)			
<p>本論文の掲げる最終的な目標は、Web上の大量のテキストの内容を統合・比較・要約し、物語として読者に提示することである。あるトピックに関連する一連の出来事を物語として捉え、登場人物、場所、出来事、そしてそれらの間の関係性などを時間軸上で表現したものをストーリーラインと呼ぶ。ストーリーラインの核となる構造は、出来事の時間軸への対応付けである。本論文では、テキスト中の時間表現を解析し時間軸に対応付ける手法、事象表現を時間軸に対応付けた言語資源 (コーパス) の構築、ある人物や製品に関する出来事を時間軸上に配置した「タイムライン」の生成手法を提案している。論文は、時間表現解析、時間情報コーパスの構築、事象に関する時間情報解析について各1章を割り当て、序論と結論の章を合わせて全5章から構成されている。</p> <p>第1章は序論である。まず、ストーリーラインとタイムラインの概略を述べたあと、従来研究のアプローチについて説明している。次に、自然言語処理分野における時間情報解釈に関する3つのアプローチについて述べ、事象情報を時間軸に直接結びつけるアプローチが、複数文書の情報の統合・比較・要約という観点において有効であることを論じている。</p> <p>第2章では、時間表現をテキストから検出し正規化する手法を提案している。時間表現の正規化とは、その表現のもつ時間情報を定められた形式に変換することである。時間表現の語順はしばしば入れ替わり多様な表現が存在しているが、既往研究では時間表現のパターンをあらかじめ用意して正規化を行っていたため、これらの表現に十分対処できないという問題があった。提案手法では時間表現の構成性に注目する。時間表現に含まれる基本的な語彙に対して時間情報を付与した語彙ルールを用意し、これを組み合わせることで多様な語順や並列構造に対処する。提案手法はニューラルネットワークを用いたモデルであり、まず系列ラベリング技術を用いて時間表現を検出したあと、検出した表現に含まれる語彙ルールを組み合わせることで正規化を行う。評価型ワークショップのデータセットを用いた実験を行い、先行研究と比較して高い精度で時間表現の検出と正規化ができることを示した。</p> <p>第3章は、事象を時間軸に対応付ける時間情報コーパスの構築について述べている。タイムラインの学習や評価を行うためには、事象を時間軸に結びつけたタグ付きデータが必要である。本研究で設計したタグ付け基準は従来研究と比較して2つの特徴をもつ。1つは幅広い表現をタグ付け対象としたことである。従来研究では一時性の強い表現を対象としてタグ付けを行っていたが、本研究では一時性の弱い表現の時間情報も重視し、これらをタグ付け対象に含めている。もう1つは、従来扱わなかった、頻度や期間などの多様な時間情報を扱うためのタグを新たに導入したことである。本タグ付け基準を用いて、日本語新聞コーパスに対してタグ付けを行った。付与されたタグの約25%が新たに導入されたタグであった。</p> <p>第4章は、事象に関する3つの時間情報解析について述べている。まず、事象間の時間的関係を推定する、外部知識を用いたニューラルネットワークモデルを提案している。本モデルは評価型ワークショップにおいて、他のチームを上回る精度を得た。し</p>			

かし、本モデルは2つの事象間の関係性に着目する一方で、テキスト中の時間情報を利用できていないという課題がある。そこで次に、事象のもつ時間情報を分析するため、3つの多値分類タスクを設計した。具体的には、事象の時間性の有無を判定するタスク、事象の時間的長さを分類するタスク、事象の発生時期を分類するタスクである。各タスクに対して、テキスト中の時間情報や文脈を考慮して分類を行うニューラルネットワークモデルを用意し、第3章で構築したコーパスを用いて実験を行った。時間性の判定は9割の精度で行えたのに対し、長さや時期の分類は5、6割の精度に留まった。最後に、一時性の強い事象を対象としたタイムライン生成手法を提案している。提案手法は従来研究と比べて幅広い文脈を考慮するもので、2段階の機械学習からなる。第1段階では局所的な情報に基づいて事象を時間軸に対応付け、第2段階では大域的な情報を用いてこれを修正する。評価型ワークショップのデータセットを用いた実験の結果、既往研究よりも高精度なタイムラインを生成できること、2段階目の処理が有効であることが示されている。

第5章は結論である。本論文を総括し、今後の課題について述べている。

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し

審査結果の要旨は日本語500～2,000字程度で作成すること。

(続紙 2)

(論文審査の結果の要旨)

本論文は、テキストの時間情報解析について、事象表現を時間軸に対応付ける手法・コーパス構築の研究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. 時間表現はテキストの時間情報の核となる情報である。従来の時間表現検出・正規化手法は、予め定めた表現パターンしか扱うことができなかった。しかし、実テキストでは時間表現の語順はしばしば入れ替わり、多様な表現が存在している。本論文では時間表現の構成性に着目し、語順に頑健なニューラルネットワークモデルを提案している。評価型ワークショップのデータセットを用いた実験の結果、提案手法が従来手法より高い精度で検出・正規化できることを示した。

2. 事象を時間軸に対応付ける時間情報コーパスは、これまで一時性の強い事象の開始時と終了時に着目して構築されてきた。本論文では、より幅広い事象表現を対象に、より多様な時間情報を表現できるタグ付け基準を設計し、この基準に基づいた時間情報コーパスを構築している。本コーパスには既に述語項関係や共参照関係がタグ付けされており、本タグ付けと合わせてテキスト中の事象・エンティティ・時間を対象とした統合的な時間情報解析に活用することが可能となった。本コーパスは今後研究者コミュニティ向けに公開される予定であり、今後の時間情報解析研究の発展に資すると考えられる。

3. 従来のタイムライン生成モデルでは、事象表現の周辺の文脈にはほとんど注意が払われてこなかった。本研究では、局所的・大局的な文脈を考慮して、事象表現と時間情報を結びつけるモデルを提案している。提案モデルは2段階の機械学習からなっており、まず局所的な情報を用いて各事象表現を時間情報に対応付け、その後大域的な情報を用いてこれを修正する。本論文では、評価型ワークショップのデータセットを用いた実験とその結果について述べられており、提案手法が従来手法より優れた手法であることが示されている。

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、平成31年2月19日実施した論文内容とそれに関連した試問の結果合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。更に、試問の結果の要旨(例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」)を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。  
要旨公開可能日： 年 月 日以降