

論文内容の要旨

博士論文題目 Constructing a Temporal Relation Identification System of Chinese based on Dependency Structure Analysis
(依存構造に基づく中国語事象表現の時間関係同定システムの構築に関する研究)

氏名 鄭育昌

(論文内容の要旨)

時間情報処理は時間表現や事象表現を同定し、表現間の時間関係を解明する研究である。言語学、心理学においては言語の時間表現の論理モデルに関する研究は数多く提案されている。しかし、時間情報の自動処理を行うため必要な時間情報タグ付きコーパスの構築に関する研究が少ない。特に、英語ではタグ付き基準 TimeML が提案されているのに対して、中国語の時間情報タグ付きコーパスに関する研究はほとんどない。本論文では、中国語の事象間の時間関係をタグづけする基準を策定し、その関係を同定する手法を提案する。

中国語は、テンスとアスペクトを陽に示す表現がなく、時間情報は語彙と文脈に依存するという特徴がある。よって、中国語の時間関係処理に対して時間表現を中心とする手法は不十分であり、事象間の時間関係同定が重要なタスクである。本論文は、中国語の事象間の時間関係を同定する手法を提案し、タグ付きコーパスを構築し機械学習器で事象間の時間関係の自動同定システムを提案する。

本論文で提案する中国語の時間関係コーパスのタグ付け基準は TimeML の基準に基づいている。提案するタグ付け基準は依存構造解析結果を用いて、タグ付け作業量を減らすことを目標にした。依存構造を用いて三つの時間関係(隣接関係、依存構造上の親子関係および兄弟関係)を定義する。さらに長距離の関係を得るために推論規則を使用して時間関係を拡張する。同一記事中ある n 個の事象に対して、すべての組合せの時間関係を考慮するため $n(n-1)/2$ 回の判定が必要であるが、提案手法による作業量は $3n$ 回の判定でよく、人手による動詞事象対の時間関係のタグ付け作業量を軽減できることを示す。

構文解析済みのコーパス Penn Chinese Treebank を基本データとして、動詞間時間関係のタグ付きコーパスを構築し、本タグ付きコーパスを用いて時間関係の自動同定システムを実装する。依存構造の情報を獲得するため、機械学習に基づく中国語の依存構造解析器を開発し、既存の中国語依存構造解析器を上回る単語の依存関係の正解率(88%)を達成した。

提案手法で構築したタグ付きコーパスを用い、機械学習器として Support Vector Machine を用いることにより、時間関係の自動同定システムを作成した。評価実験において提案した各種類の事象の時間関係の解析正解率は 68%-70% に達し、記事中のすべての可能な事象間の時間関係の 52% が提案手法によるシステムで同定できることを示す。

氏名	鄭育昌
----	-----

(論文審査結果の要旨)

平成19年12月27日に開催した公聴会の結果を参考に平成20年2月18日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

鄭育昌は、本博士論文において、中国語文書中の動詞が表す事象間の時間関係解析を目的とし、時間関係タグ付けを行うための効率的な手法、および、機械学習に基づく事象間の時間関係同定を行う手法を提案した。特に係り受けに基づく統語解析からえられる情報がこれらのタスクに重要な情報を提供することを示し、自ら中国語の単語係り受け解析器の構築も行った。本論文の主な成果は次の通りである。

1. 中国語テキストへ時間関係のタグ付けを効率化するために、次のような方法論を示した。すなわち、依存構造によって得られる三つの関係(隣接関係, 親子関係, 兄弟関係)に対してのみタグ付けを行い、それ以外の関係にある事象間の時間関係を推論規則によって拡張する方法である。これにより、人手による動詞事象対の時間関係のタグ付け作業量を大きく軽減できることを示した。
2. 構文解析済みのコーパスを学習データとし、機械学習に基づく中国語の依存構造解析器を開発し、既存の中国語依存構造解析器を上回る単語の依存関係の正解率を達成した。
3. 提案手法で構築したタグ付きコーパスを訓練データとし、Support Vector Machinesを用いて、時間関係の自動同定システムを作成した。評価実験において提案した各種類の事象の時間関係の正解率が68%-70%と、高い性能を実現できることを示した。

中国語の文書中の事象の時間関係解析のための基礎データの作成、時間関係抽出に有用な機械学習に基づく中国語係り受け解析手法の提案、および、事象間の時間関係の自動同定手法を提案した本研究は、独創性が高く、しかも実用的であり、自然言語処理の分野において高い貢献があると評価する。

よって、本論文は、博士(工学)の学位論文として価値あるものと認める。