論文内容の要旨

博士論文題目　Extracting Named Entity Relations from Large Text Corpora
(大規模テキストからの固有名詞間の関係抽出)

氏名　　平野 徹

(論文内容の要旨)

Web 上に存在する膨大なテキストは広い分野をカバーしており，巨大な知識源と考えることができる．テキストを知識源として活用するには，個々のテキストに含まれる情報を抽出し構造化された形式に変換する必要がある．本論文では，情報検索や質問応答などのアプリケーションにおいて重要な知識源となる実世界の実体を指し示す固有名詞間の関係情報を抽出することを目的とする．

本研究で抽出する関係情報は，個々の文書 (D) で言及されている意味的な関係のある固有名詞の組 (X,Y) とその間の関係 (R) を［X, Y, R, D］の構造化された形で表現した情報である．例えば,「浅田真央の姉の舞は…」の文書 (ID = 001) から抽出される関係情報は［浅田真央，舞，姉，001］となる．この例では，固有名詞間の関係を表す表現「姉」が文書中に明記されているが，明記されていない関係情報も本研究では抽出対象とする．例えば,「民主党の鳩山氏は…」の文書 (ID = 002) からは，関係を示す表現が文書中に明記されていないが「民主党」と「鳩山」の間に「党員」の関係があると読み取れるため，関係情報［民主党，鳩山，党員，002］が抽出できる．

上記の 2 種類の関係情報を抽出するために，本研究では，(1) 入力文書内で共起する固有名詞の組から何らかの関係を有する組を選択し（関係性判定），(2) 選択された組がどういう関係にあるのかを示す表現を入力文書から抽出（関係表現同定）し，(3) 関係を示す表現が文書中に存在しない組の関係を推定する（関係推定），3 つのタスクに分ける．(1) 関係性判定タスクは，従来，同一文内で共起する組に対して関係の有無を判定することはできたが，日本語において頻出する文をまたいで共起する組に対して判定することはできなかった．そこで，照応解析で用いられている文脈的情報を関係性判定タスクに適用し，文をまたいで共起する組に対しても判定可能な手法を提案した．(2) 関係表現同定タスクでは，文の構造情報だけでは関係表現か判断できない事例に対して，大規模テキストから自動獲得した 2 種類の外部知識を利用する手法を提案した．1 つは関係を示す名詞を推定した語彙的知識で，もう 1 つは対象組の過去の関係から現在の関係を予測する関係予測モデルである．評価実験では提案した 2 種類の外部知識を用いることの有効性を確認した．(3) 関係推定タスクにおいては，関係情報の類似性に基づく推定手法の根幹を担う類似度尺度について、上記 (1)(2) のタスクで自動獲得された関係情報に基づく従来の固有表現組 (X,Y) の類似度と固有表現組の出力する文書 (D) の類似度を組み合わせた尺度を提案した．評価実験では 2 種類の類似度の混合割合を変えた実験を実施し，提案手法の有効性を確認した．

氏 名	平野 徹

（論文審査結果の要旨）

平成24年7月24日に開催した公聴会の結果を参考に平成24年9月5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．

平野 徹は，本博士論文において，大規模なテキストデータから，固有名詞間の関係情報を抽出する手法を提案した．従来行われている類似研究では，固有表現の関係を1つあるいは少数の限定された関係を対象にしているか，あるいは，関係名が明示的に表現されたテキストからの抽出を対象とすることが多かった．本論文では，固有名詞間の関係を限定せず，かつ，関係名が具体的に示されていない固有名詞間の関係を自動推定するという，従来研究に比べて遥かに広い範囲を対象にする方法を提案した．本研究の特徴は次のようにまとめることができる．

1. テキスト中に存在する2つの固有名詞の関係が，テキスト中に明示的に現れる場合だけでなく，明記されていない場合をも対象に，固有名詞間の関係抽出手法を提案したこと．

2. テキスト内の固有名詞として，1文中やある特定の表現で結ばれた狭い範囲に出現する組だけを対象とするのではなく，照応解析の手法を利用して，文をまたいで出現する固有名詞の組からも関係抽出が可能であることを示したこと．

3. 固有名詞間の関係推定を行うために，関係を表現する単語の特徴に関する考察を行い，その特徴を明らかにする客観的な方法を提案したこと．また，特定の固有名詞の間の関係の時間的な変化を学習することにより，過去の関係を利用し，現在の関係を推定する手法を提案したこと．

4. 固有名詞間に明示的に示されていない関係抽出のため，固有名詞の組の類似性に加えて，それらの組が出現する文書の類似性を利用した手法を提案し，その有効性を確認したこと．

大規模テキストデータに出現する固有名詞の間の意味関係抽出のための種々の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．