人名と実体名間の関係の推定手法に関する研究

堂前 友貴

本研究では，特定の人名についての情報の整理を支援することを目的とし，人名と，人名と一緒に出現しやすい実体名との間の関係のタイプを推定する研究に取り組んだ。人名が未知のものであったとしても，実体名との間の関係を推定し，提示できれば，人物についての理解を支援できる。本研究では，人名と実体名（“ 神田はるか ”, “ 東京 ”）を入力し，あらかじめ定義した関係（“ 人-著作 ”）への推定を行う関係推定器の構築に取り組んだ。

人名と周囲に現れる実体名との間に成立する関係は，社会においてよく利用される情報という観点から指定しておくことができる。そこで，本研究では，人名と周囲に現れる実体名を分析し，推定する関係の定義を行った。関係の分析には拡張固有表現タグ付きコーパスを使用し，人名と同一文中に多く現れる拡張固有表現を分析することで，上記の観点から 10 種類の関係を定義した。

このように定義した関係にもとづき，単語対を入力とし，関係を推定する関係推定器を構築した。推定を行うためには何かしらの手がかりが必要となるため，単語対を含む文を Web 上のテキストから収集し，推定を行う。処理の流れは，次の通りである。(1) 単語対を含む文を Web 上のテキストから収集し，(2) 1 文ずつ関係のタイプを推定し，(3) 関係のタイプの多数決で，出力する関係ラベルを決定する。

この処理では，1 文ずつ関係のタイプを推定することが重要となる。本研究では，1 文ずつの関係のタイプの推定に分類器を使用した。分類器は機械学習の一手法であり，何かしらのラベルが付与されたデータが必要となる。ラベル付きデータの取得は，大規模なものを人手で作成することはコストが高いため，最近では，初期データから学習データを自動的に拡張する研究が活発に行われている。これは，ある関係が成立する単語対である関係例を収集し，その単語対を含む文をさらに収集するものである。本研究もこの方式を利用し，関係例を取得する際の手がかりの制約として，拡張固有表現を採用した。ラベル付きデータの取得は，次の通りである。(1) 拡張固有表現タグ付きコーパスを分析し，関係例を取得するためのパターンを定義し，(2) コーパス上で関係例の取得し，(3) 関係例を含む文を Web 上のテキストから取得する。また，分類を行う際に使用する特徴量（素性）には，(A) 実体名の品詞，(B) 人名と実体名のどちらが先にくるか，(C–F) 単語対のそれぞれの前後に現れる形態素と (G–J) その品詞　の 10 要素を用いた。

分類器の性能を，人手判定したデータで評価した結果は，精度 0.86，再現率 0.83，F 値 0.84 であった。また，構築した関係推定器に対し，各関係 50 組，関係が成立しないもの 50 組の計 550 組の単語対で評価を行ったところ，精度 0.71，再現率 0.76，F 値 0.73 となった。今後の課題としては，素性の改良による分類器の精度向上や，推定する関係数の拡張などが挙げられる。

（指導教員　関 洋平）