

平成27年度 情報工学コース卒業研究報告要旨

| | | |
|---|---------------------------------|---------|
| 外山 研究室 | 氏 名 | 山 腰 貴 大 |
| 卒業研究題目 | 語彙の意味と機能の類似性に着目した法令文の 並列構造解析 | |
| <p>法令には人々が守るべき義務や得られる権利が記されている。法令を読み、正しく理解することで多くの社会的利益を得ることができる。しかし、法令文は一般の人々にとって非常に読みにくいものである。その原因として、法令文特有の言い回しや専門用語、複雑な並列構造が頻出することが挙げられる。そのため、法令の理解を助けることを目的に、法令用語のシソーラスを構築する研究や法令文を読みやすく表示する研究が行われている。これらの研究に必要不可欠な基盤技術の一つとして、並列構造解析がある。</p> <p>一方、法令文の並列構造は、解釈の曖昧性を排除するため、明確な規則に従って記述されている。その規則に基づいて決定的に並列構造を同定する手法（以下、決定的手法）がこれまでに提案されているが、必ずしも十分な精度を達成していない。それに対して、法令文における並列構造に関する記述規則を文脈自由文法（CFG）で形式化し、そのCFGから導出される全候補の中から最適な並列構造を同定するアプローチが検討されている。水尻は、このアプローチの一步として、法令文の並列構造解析のためのCFGを構築し、そのCFGにより、実データにおける約90%の並列構造に対して正しい構文木を導出できることを確認している。しかし、この研究では、オラクル付きのボトムアップチャートパーザによりCFGのカバー率を検証しているにすぎず、並列構造解析を実現しているわけではない。さらに、このCFGには、法令文の記述規則に反する並列構造を生成する可能性があるなど、一部に不完全な点がある。</p> <p>そこで本研究では、上述したCFGによるアプローチに基づいて、法令文の並列構造解析を実現する。具体的には、上述のCFGを改良するとともに、CFGから導出される候補の中から適切な候補を自動的に選択するためのスコアリング手法を提案する。改良したCFGには、どのように生成規則を適用しても不適切な並列構造を生成しないように、より強い制約を加えている。提案するスコアリング手法は、CFGにより導出された並列構造を表す部分木に対して、並列構造としての妥当性を示すスコアを付与する。このスコアは並列構造で列挙されている語句の類似性に基づいて算出される。本手法では、類似性の尺度として意味的類似性と機能的類似性を導入する。意味的類似性は、文節に含まれる形態素の意味の近さを表す。例えば、形態素「りんご」は、野菜の一種の「白菜」よりも、同じ果物の一種の「みかん」の方が意味的に類似していると考ええる。機能的類似性は、文節の文における役割の近さを表す。例えば、文節「見る」は、目的格である「視聴を」よりも、同じ動詞である「読む」の方が機能的に類似していると考ええる。</p> <p>本手法を評価するために、2法令592文（括弧付き表現は別文として数える）を用いて並列構造解析の比較実験を実施した。まず、意味的類似性と機能的類似性の効果を確認するために、類似度として意味的類似性のみ、機能的類似性のみ、その両方を用いた場合で解析した。その結果、それぞれの場合の適合率は45.5%(=304/668)、48.7%(=325/668)、52.5%(=351/668)となり、両方を用いる場合が最も高く、その効果を確認した。次に、水尻が構築したCFGとの比較実験を行った。その結果、改良したCFGを用いたときの適合率52.5%は、水尻のCFGを用いたときの適合率49.3%(=331/672)を上回っており、改良したCFGの有効性が確認できた。最後に、先行研究の決定的手法との比較実験を行った。その結果、本手法の適合率52.5%は、決定的手法の適合率35.6%(=276/775)を上回っており、本手法の有効性が確認できた。</p> | | |