

審査の結果の要旨

氏名 山田 恵美子

本研究は、コンピュータによる医学テキストの処理の高度化において重要な役割を果たすと考えられる、複合語としての医学専門用語、特に疾患名について、その内部構造の表現方法を提案したものであり、下記の結果を得ている。

1. 医学専門用語の内部構造を、文字を単位語とした係り受け構造によって表現することを提案した。係り受け構造の決定方法は、前方・後方探索による語の分割位置決定と、分割前の語との関係として **is-a** 関係または **part-of** 関係にある構成要素を係り先とする係り受け関係の決定の2段階から成るものであった。更に、省略や縮退によって複合語には現れないが構成要素には現れる文字の位置を複合語の内部構造として明示的に示すため、上述の係り受け構造において通常の係り受け関係の他に文字を生成する係り受け関係を導入し、医学専門用語の内部構造をラベル付き係り受け構造として表現することを提案した。
2. 内部構造を用いた、疾患名の自動 ICD コーディング方法を提案した。提案手法は ICD 分類の見出し文字列からコーディングルールを作成し、入力疾患名の内部構造にルールを適用するものであった。実験の結果、提案手法のコーディング精度は比較手法よりも高いことが示された。
3. 文字列として表現された医学専門用語を自動解析し、1 で提案した内部構造を出力する解析器を、既存の汎用係り受け解析器 **MaltParser** を用いて作成した。実験の結果、83.7%の精度で語を解析可能であることが示された。

以上、本論文は医学専門用語の内部構造の表現方法を提案し、これを用いることで既存の自動 ICD コーディング手法の精度が向上することで内部構造の医療における有用性を示し、また通常の文字列表記から内部構造表現を得る自動解析が有望な精度で可能であることを示した。さらに本研究はこれまで対象とされてこなかった複合語を構成する文字の省略・縮退を扱う枠組みを初めて提案したのもでもあり、今後の医療情報処理におけるテキスト解析の発展に重要な貢献をなすと考えられ、学位の授与に値するものと考えられる。