

栗田修平（黒橋禎夫教授）

「Neural Approaches for Syntactic and Semantic Analysis」

（構文・意味解析に対するニューラルネットワークを利用した手法）

平成 31 年 3 月 25 日授与

人間の言語を理解し、コミュニケーションを図る事ができるシステムを構築する際に、自然言語処理は本質的な技術となる。自然言語処理ではまず統語的ないし意味的な手法で入力文を解析する。これらのテキスト解析手法は、自然言語処理の基礎解析をなし、テキスト入力を利用する多くの現代的なシステムにとって必要な解析となる。しかしながら、自然言語入力文を解析するモデルを開発するためには、以下のような理由から困難である。第一に、文を解析するモデルは多くの場合、他の解析手法と同時に使用される。他のモデルから出力された結果には、しばしばバイアス（偏り）や誤りが含まれ、これをそのまま解析に使用することは難しい。第二に、構文的ないし意味的な解析は、しばしば、既存のコーパスからは獲得が難しい外部知識を必要とする。第三に、自然言語処理の基礎解析における既存手法は、学習コーパスに対するドメイン依存性が強かった。既存手法はしばしば詳細な素性選択に依存しており、モデルの汎用性や表現力の自由度を犠牲にしていた。これらが、既存の自然言語テキストに対する基礎解析の課題であった。

過去の 10 年間で、ニューラルネットワークは機械学習を使用する多くの研究で共通して使われる基盤技術に成長した。これは、ニューラルネットワークが、その内部変数の多さゆえに強い表現能力を持っていることによる。自然言語処理においては、しかしながら、単純な教師ありニューラルネットワークを文解析の分類器として使用したのでは、必ずしも既存手法よりも高い精度を期待できない。これは、先程述べた既存手法における課題が、単純にニューラルネットワーク手法を適用するだけでは解決できないからである。これらの問題を解決するため、本論文では、自然言語処理の基礎解析において、特に中国語の統合構文解析、日本語省略解析、英語の意味依存構造解析に着目し、これらの諸問題に対して、統合解析ニューラルネットワークモデル、生成モデル、そして強化学習モデルを利用した全く新しい手法を提案する。中国語の構文解析では、単語分割、品詞タグ付けおよび単語の係り受け解析を統合して行うモデルの探索を行った。日本語省略解析では、大規模なラベルなしコーパスから敵対的生成ニューラルネットワークを応用して、省略されやすい述語項に関する知識抽出を行った。英語の意味依存構造解析では、強化学習を利用して、英語の文中で解析を行う箇所を自ら決定するモデルを提案した。

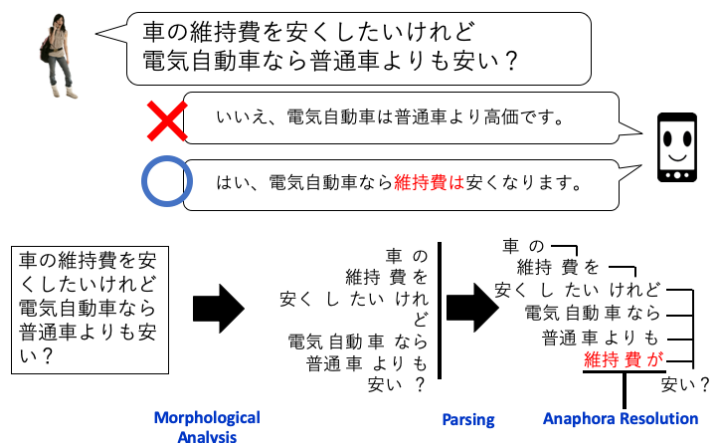


図 1 自然言語処理による文解析