

論文の内容の要旨

論文題目 単語間意味関係知識の獲得と応用

氏名 鷺尾光樹

単語間意味関係とは、単語・概念間にある関係のことである。これには、同義関係 (*car-auto*)、上位下位関係 (*animal-cat*) などが含まれる。これらの知識は自然言語理解において重要な要素であり、含意関係認識や質問応答などの高度な意味処理を要するタスクで用いられる。

自然言語処理における単語間意味関係知識は主に、WordNet に代表される語彙知識ベースに蓄えられ、様々なタスクで利用されるが、人手で作成される語彙知識ベースはカバーされているドメインも限られており、拡張にも大きなコストがかかる。その結果、語彙知識ベースを用いた自然言語処理システムが、未知語や新語に対応できないといった問題が生じてしまう。この問題を解決するために、コーパスから単語ペアの意味関係知識を自動的に獲得する技術が研究されている。また、近年ではニューラルネットワークモデルが様々な自然言語処理タスクで性能を向上させており、ニューラルネットワークに知識を挿入する場合は、低次元で密なベクトルの形式で知識を表現するのが有力なアプローチである。よって、コーパスから獲得された知識は、適切に単語間意味関係を表現しており、さ

らに知識の表現は後続のタスクで扱われる埋め込みの形式で獲得されることが望ましい。しかし、既存の獲得法は、意味関係表現としての適切さと表現できる単語ペアの範囲の広さが両立していないという問題がある。また、単語間意味関係知識は高度な意味処理に必要であるが、このような知識をどのようにシステムに組み込むかは自明ではなく、獲得された意味関係知識を新たなタスクにどのように応用し、単語間意味関係知識の応用の幅を広げるか、それ自体も問題である。

本研究ではこれらの問題を解決するために、二つの研究に取り組んだ。一つ目の研究は、コーパスから意味関係を自動的に埋め込み形式で獲得する、関係パターンを用いた単語ペア埋め込みの教師なし学習である。関係パターンとは文中で共起した二語を結びつける単語系列、あるいは依存構造パスであり、単語間の意味関係を表現する上で重要な特徴である。従来手法では、関係パターンを用いた場合、埋め込み表現を獲得できる単語ペアが実際にコーパス上で共起したものに限定されてしまうという大きな制約があった。本研究ではニューラルネットワークの汎化能力を用いることで、単語埋め込みが割り当てられている任意の単語ペアについて、関係パターンの情報を捉えた単語ペア埋め込みを獲得することができる手法を提案する。提案する単語ペア埋め込みの学習法の評価は、単語ペアの単語間意味関係類似度の評価ベンチマークと、意味関係識別のデータセットで行い、有効性を確認した。

二つ目の研究では、単語間意味関係知識の新たな応用として、単語ペア埋め込みを定義文処理に応用する手法を提案する。定義文処理は、定義文からの単語埋め込みの獲得や、単語埋め込みからの定義文の生成を含む、定義される見出し語と定義文間のマッピングをモデリングするタスクである。定義文には見出し語と意味関係を持つ単語が存在し、これらをモデリングに組み込むことで、より適切な定義文処理が行えることを示した。

以上の二つの研究を通して、本博士論文では、コーパスから単語間意味関係知識を自動獲得し、それを自然言語処理に活かすことの重要性を示す。