

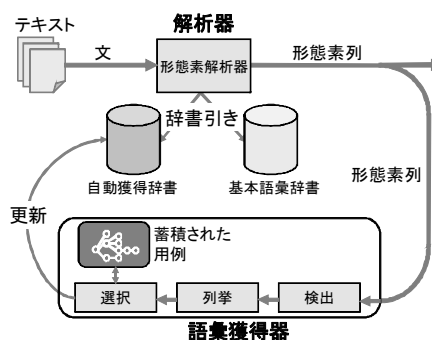
タイトル

自然言語の解析のためのテキストからの語彙の自動獲得

Automatic Lexicon Acquisition from Text for Natural Language Analysis

概要

辞書にない形態素(未知語)は自然言語の解析誤りを引き起こすため、これまで人手による大規模な辞書整備が行われてきた。しかし、語彙には際限がなく、ロングテールにあたる低頻度の形態素まで人手で網羅するのは現実的でない。そこで、我々はテキストから語彙知識を自動獲得する手法を提案している。この手法により、テキストから獲得された未知語が人手の介在なしに辞書に追加され、それによりテキストの解析精度が向上する。



URL

産業界への展開例・適用分野

形態素解析はベンチマークたる新聞記事に対しては実用的な精度が報告されている。しかし、他分野のテキストに適用すると未知語が原因で精度が悪化する。しかも、専門用語など、特殊だが重要な語彙ほど誤りやすい。現在は分野適応のために人的リソースを割いて辞書整備が行われているが、提案手法はこうした作業の自動化を目的とする。

研究者

	氏名	専攻	研究室	役職(学年)
展示担当者	村脇 有吾	知能情報学	黒橋研究室	博士2年