

審査結果の要旨

論文提出者氏名 梶博行

本論文は、「Automatic Construction of Word Sense Association Networks」(語義関連ネットワークの自動生成)と題し、6章からなる。自然言語処理システムでは、語の文法的・意味的な知識を記述した高精度の辞書が重要な役割を果たすが、言語現象の多様性・個別性および語彙の大規模性のために辞書の作成が隘路となっていた。本論文では、2つの異なる自然言語にまたがる2言語コーパスを利用して、自然言語の意味処理の基盤となる「語義関連ネットワーク」を自動生成する手法を提案し、そのために必要な技術を論じたものである。

第1章「Introduction(緒言)」では、本研究の背景と目的を述べるとともに、関連分野である意味辞書、コーパスからの対訳語の抽出、コーパスに基づく語義の曖昧性解消、およびコーパス中の分布に基づく語のクラスタリングに関する従来研究を整理している。

第2章「From Word Associations to Word Sense Associations: An Approach Using Bilingual Corpora(語関連から語義関連へ: 2言語コーパスを用いる1アプローチ)」では、語義関連ネットワーク(節点が2言語の同義語集合で定義される語義を表し、枝が語義間のトピック的な連想関係を表すグラフ)という概念を導入した上で、語義関連ネットワーク自動生成手法を概観している。この語義関連ネットワーク自動生成手法では、2言語コンパラブルコーパスから各言語の語関連ネットワークを生成し、対訳辞書を利用してそれらを反復的に対応付けることによって、語義関連ネットワークを生成する。この手法の特色は、語の語義への分割と語関連の語義関連への変換の困難さを緩和するために、語義関連ネットワークの中間形式として、個々の語に対する語義リストと語義-手がかり語相関行列を設定し、与えられた語義リストに対し語義-手がかり語相関行列を計算する「語義-手がかり語の相関計算」と、語義-手がかり語相関行列を利用して語義リストを生成する「訳語集合のクラスタリング」を交互に実行する方式を採用したことである。本手法と相補的な関係にある概念分類型の意味辞書 WordNet との結合可能性についても議論している。

第3章「Extraction of Translation Equivalents Based on Contextual Similarity(文脈類似度に基づく対訳語の抽出)」では、2言語コーパスから対訳語ペアを抽出するために、両言語の語を共起頻度付きの共起語集合で特徴付け、基本語の対訳辞書を用いて共起語集合の類似度を計算し、類似

度の高い語のペアを抽出する方法を提案している。この方法は、統計的方法と言語的方法の長所を併せ持ち、多様なタイプの対訳語ペアや小さなコーパスから低頻度の対訳語ペアの抽出ができるなどの利点がある。本方法の有効性は実験的に確認されている。

第4章「Interactive Calculation of Sense-vs-Clue Correlations Based on Translingual Alignment of Word Associations(語関連の言語間対応付けに基づく語義 - 手がかり語相関の反復計算)」では、「語関連の対応付けの確度は当該語関連に随伴する関連語の間の対応付けの確度に依存する」、「語義と手がかり語の相関は当該語義と随伴する手がかり語との相関に依存する」という2つの仮説に基づいて、語義 - 手がかり語の相関をコンパラブルコーパスに基づいて反復計算する方法を提案している。この反復計算はスパースなデータに対するスムージング効果をもち、コーパス中のインスタンスの数が比較的小さな多義語にも適用可能であり、収束が速く、計算量の面でも十分実用的であるという利点がある。評価実験により、有効性が実証されている。

第5章「Clustering of Translation Equivalent Patterns Based on Similarity of Translingual Distributed Patterns(原言語での分布パターンの類似度に基づく訳語のクラスタリング)」では、第4章の語義 - 手がかり語相関計算方法を用いて原言語に写像した分布パターンの類似度に基づいて、多義語に対する訳語の集合を階層的にクラスタリングする方法について述べている。この方法は、類似度を計算する相手の訳語あるいは訳語集合を除外して計算した語義 - 手がかり語相関行列を併用しているため、訳語の使用頻度の差の影響を小さくすることができる。本手法は、コーパス中で使用されている語義のみを高精度で獲得することができるとともに、計算量の面でも実用的である。評価実験によりこの方法の有効性が確認されている。

最後に、第6章「Conclusion(結言)」では、本研究の総括を行い、併せて将来の展望について述べている。

以上を要するに、本論文では、今後の自然言語処理の基盤となる意味辞書を構築するために、文脈類似度に基づく対訳語の抽出、コンパラブルコーパスに基づく語義 - 手がかり語の相関の反復計算、原言語に写像した分布パターンの類似度に基づく多義語に対する訳語集合の階層的なクラスタリングによって、2言語コンパラブルコーパスから語義関連ネットワークを自動生成する手法を提案し、評価することによってその有効性を示したものであり、電子情報工学上貢献するところが少なくない。

よって本論文は博士(工学)の学位請求論文として合格と認められる。