

学位論文題名

帰納的学習による表層文と意味表現間の
変換規則獲得に関する研究

学位論文内容の要旨

人間の知的活動におけるもっとも優れた能力の一つに言語の使用が挙げられる。言語の使用によって、人類は深遠なコミュニケーションを可能とし、知識や文化を広く伝搬することに成功してきた。それゆえ、20世紀半ばに電子計算機が発明されて以来、言語を操るシステムを作ることが大きな目標となった。しかし、50年近い集中的な研究にも関わらず、いまだ人間のように言語を操るシステムは実現していない。自然言語の顕著な特徴として、表面的には極めて曖昧で厳密性に欠けるように見えることがあげられる。人間同士の会話では、話し手と聞き手は両者の共有する社会的、文脈的知識を活用し、表面上の曖昧性を解消しながら理解を進めている。しかし、計算機にこうした曖昧性解消のメカニズムを実現することは極めて困難であり、自然言語処理はこの曖昧性との闘いであったといえる。

自然言語処理の伝統的な手法では、形態素解析、構文解析、意味解析などの解析過程によって表層文の持つ意味を定式化して意味表現に変換し、これを言語理解部への入力とする方法が多く用いられてきた。これはこの解析、変換過程を通して言語の持つ曖昧性や多義性を解消することを目的としている。多様な言語表現を解析するためには、それぞれの解析過程における規則系の充実が欠かせないが、前述のような自然言語の持つ曖昧性や言語表現の非規則的な多様性が障害となり、無矛盾性を保ちながら規則系を拡大することは容易ではない。すなわち、伝統的解析手法の問題点は膨大な言語規則をいかに構成し、管理するかにあるといえる。規則を単純に増やすことは問題の解決にならず、それぞれの規則が適切に適用され得る規則系を構築する手法が求められる。

本論文では、計算機による深い自然言語解析のための規則系の構築に帰納的学習を導入する手法を提案する。本論文における帰納的学習とは、あらかじめ与えられた表層文と意味表現の対の集合を、既知の規則の再帰的な適用や部分規則の推定によって一般化し、表層文から意味表現への変換規則を自動的に獲得することをいう。本論文で提案する手法は人間の言語獲得過程を参考にしている。人間の子供は母国語をおおよそ2歳から6歳までの短い期間に習得することができ、その後も日常の言語使用を通じて新しい表現や語彙を獲得し、言語知識を維持向上して行く。提案手法は、計算機上にこうした能力を実現することを目標としている。

本手法では、獲得された規則やその優先度が動的に変化し、状況に応じて適応的に成

長する。それゆえ、人手で規則を管理する必要がなく、規則管理者の主観を排除でき、また、規則間の優先度が学習を通して自動的に調節されるため、高品質な出力を得ることができる。一般に自然言語処理システムでは、対象とする分野に合わせて規則のチューニングを行うことが普通であるが、本手法では継続的学習によって規則を自動的に対象分野に適応させることが可能である。

本論文は6章からなる。第1章は序論で、自然言語処理の現状と問題点を概観し、提案手法の概要を説明する。

第2章では、二部に分けて意味表現の定義を行う。第一部では単語の意味に対応する概念を定義する。単語は多義性を持つことが多く、これらの語義を分類し定義する必要がある。本論文では異言語対訳を用いた概念定義を提案する。前述のように規則選択の曖昧性の解消のためには文脈、一般知識などさまざまな情報が必要であるが、提案手法では概念間の相互関係を与える異言語シソーラスから作成した情報を一般知識として与えている。第二部では表層文中の各概念の相互関係を定式化する。本論文では深層格フレームを用いた意味表現定義を採用している。

第3章では、知識を計算機上に実現するモデルを示す。ここでは大量の知識の相互関係の表現方法と、それぞれの知識の優先度を局面に応じていかに設定するかがポイントとなるが、本論文では単一の重み付きリンク体系上に全てのデータを表現する手法を提案する。表層文の一部や単語概念、シソーラス、意味表現などのデータは重み付きのリンクで結ばれ、ネットワーク構造によって規則系を構成する。リンクは各知識の相互関係を示し、重みはそのリンクの優先度を示す。リンク表現された規則には規則を想起するために想起リンクが張られる。想起リンクのソースノードには、単語、シソーラス、上位の規則を想起するための規則片などが含まれる。

第4章では、変換規則の獲得手法および適用手法について述べる。本論文で提案する規則の獲得アルゴリズムは帰納的学習であり、既知規則の適用と規則推定によって入力データを一般化し、規則として獲得する。規則獲得を通じて、リンクは有効な規則を想起するように張られる。リンクの重みは、想起した規則の適用の正否に応じて規則獲得過程で自動的に設定され、適応的に変化する。規則の適用プロセスでは、入力文に対し適用規則を選択し意味表現を生成する。規則は重みの大きい想起リンクをたどることで直接に想起されるため、高速な規則適用が可能である。規則適用は先頭単語から後続の単語へと、想起中の規則の妥当性を検証する形で進み、システムは妥当性が低いと判断するとバックトラックによって新しい規則を想起する。妥当性検証の過程ではシソーラスデータが参照される。さらに本章では本手法を用いた規則獲得、適用の実験システムを構築し、その有効性を実証した実験結果を示す。

第5章では、本手法の応用として機械翻訳システム、質問応答システムについて述べる。機械翻訳システムについては、意味表現から表層文を導出するアルゴリズムについて述べ、簡単なプロトタイプを作成した結果についても報告する。質問応答システムについては、本手法を用いる有効性と必要な付加機能について説明し、システム設計の概要を述べる。

第6章は結論で、本論文で提案した手法をまとめ、実験システムから得られた結果からその有効性について評価する。さらに、今後の展望について述べる。

学位論文審査の要旨

主 査 教 授 梶 内 香 次
副 査 教 授 青 木 由 直
副 査 教 授 北 島 秀 夫
副 査 教 授 宮 永 喜 一

学 位 論 文 題 名

帰納的学習による表層文と意味表現間の 変換規則獲得に関する研究

人間と計算機間の自然言語による自由な情報伝達を実現することは、計算機が出現して以来の重要な研究課題であるが、未だ実現していない。その大きな理由として、言語の持つ曖昧性があげられる。人間同士の情報交換では、話し手と聞き手は両者の共有する知識を用い、曖昧性を解消して意味の理解を行っているのに対し、計算機にこのような曖昧性解消のメカニズムを実現することは極めて困難である。

計算機による自然言語処理の伝統的な手法では、形態素解析、構文解析、意味解析などの過程を通じて表層文の持つ意味を定式化して意味表現に変換し、これを言語理解部への入力する方法が多く用いられ、これらの過程を通して曖昧性や多義性の解消を図っている。このため、各々の解析課程における規則系の充実が必要であるが、無矛盾性を保ちながら規則系を拡大することは容易ではなく、各々の規則が適切に適用され得る規則系を構築する手法が求められている。

本論文は、計算機による自然言語解析のための規則系の構築に帰納的学習を導入する手法を提案したものである。この手法では、与えられた表層文と意味表現の対の集合を、既知の規則の再帰的な適用や部分規則の推定によって一般化し、表層文から意味表現への変換規則を自動的に獲得することを可能とする。また、獲得された規則やその優先度が状況に応じて適応的に変化するため、従来手法の多くに見られるように、対象分野に合わせて人手により規則のチューニングを行う必要がなく、継続的学習によって規則を自動的に対象分野に適応させることが可能である。

以下、本論文の構成を示す。

第1章では、自然言語処理の現状と問題点を概観し、提案手法の概要を述べている。

第2章では意味表現の定義を行っている。まず、単語の意味に対応する概念を定義

している。単語の多義性のため、語義を分類する必要があるが、本論文では異言語対訳を用いた概念定義法を提案し、規則選択の曖昧性を解消するために、異言語シソーラスにより概念間の相互関係を一般知識として与えている。ついで、表層文中の概念の相互関係を定式化し、深層格フレームを用いて意味表現の定義を行っている。

第3章では、知識を計算機上に実現するために、単一の重み付きリンク体系上に全てのデータを表し、知識間の相互関係を表現するとともに、それぞれの知識の優先度をリンクの重みによって表現する手法を提案している。さらに、リンク表現された規則に想起リンクを張り、単語、シソーラス、上位規則を想起する手法を示している。

第4章では、帰納的学習アルゴリズムに基づく変換規則の獲得および適用手法について提案している。本手法では既知規則の適用と推定によって入力データを一般化し、新たな規則として獲得する。ここで、有効な規則を想起するリンクが張られ、その重みは、想起した規則に応じて適応的に変化する。規則は想起の妥当性を検証する形で適用され、重みの大きい想起リンクを直接たどることにより、高速な規則適用を可能にしている。また、適用された規則の妥当性検証の過程ではシソーラスデータが参照される。本章では、以上の手法について述べるとともに、本手法を用いた規則獲得、適用の実験システムを構築し、その有効性を実証した実験結果を示している。

第5章では、本手法の機械翻訳システム、質問応答システムへの応用について述べている。機械翻訳システムについては、意味表現から表層文を導出するアルゴリズムについて述べ、実験用プロトタイプを構築して実験を行った結果について述べている。質問応答システムについては、本手法を用いることの有効性と、質問応答系として必要な付加機能について説明し、システム設計の概要を述べている。

第6章は結論で、本論文で提案した手法を総括し、実験システムから得られた結果からその有効性を評価している。

これを要するに、著者は、計算機による自然言語解析のための規則系の構築に帰納的学習を導入する手法を提案し、獲得された規則とその優先度が状況に応じて適応的に変換し、規則を対象分野に適応させることが可能であることを示し、実験によりその有効性を実証するとともに、機械翻訳システム、質問応答システムなどへの応用可能性について述べたもので、自然言語処理工学ならびに情報メディア工学の発展に寄与するところ大である。

よって著者は、北海道大学博士（工学）の学位を授与される資格あるものと認める。