

Web を情報源とした持続的な知識収集基盤の構築とその応用

研究代表者

大原 剛三

【研究の目的】

人工知能を利用した知的システムの実現には、その知的活動を支える知識の利用が不可欠である。その知識は、概念間の階層関係を定義した概念辞書やオントロジーのように事前に人の手によって整備・構造化されている場合もあれば、収集した生データから機械学習などの技術を用いて導出する場合もある。後者の場合であっても、学習に必要なデータの収集とラベル付けは人手に依るところが大きい。しかしながら、我々を取り巻く環境は日々変化しており、常に新しい概念が生まれ、新しい事象が観察されている。そのような動的に変化する環境に柔軟に対応し、持続的に活動を続ける知的システムを実現するためには、その活動の源泉となる知識をシステム自身が必要に応じて動的に収集・獲得できる必要がある。本研究では、そのような知識獲得のための情報源として Web に着目する。Web 上には、構造化された情報、構造化されていない情報、文書のような半構造化情報など、様々な情報が公開されており、これまでにその情報を特定の知的システムに活用する試みは数多く見られた。それらの既存アプローチとは異なり、本研究では、特定の知的システムを前提とするのではなく、多様な知的システムと連携可能な汎用的な持続的知識獲得フレームワークの実現を目指す。具体的には、既存の知的システムが新たな知識を要求するのは自身のもつ知識だけでは現状のタスクの遂行が困難となった場合であるとの仮定の下、1) その遂行困難となったタスクにかかる情報をトリガーとして必要な情報のみを選択的に Web から収集する**能動的情報収集システム**、2) 収集した情報を既存の知識・概念ベースに組み込む**知識構造化システム**、3) 構造化した知識・概念ベースの内容を知的システムと共有可能にする**知識共有基盤**の実現を、機械学習(大原・伊東)、知識処理(森田・谷津)、Web 情報処理(Dürst・莊司)を専門とする研究グループが協力することにより機動的に進める。最終的にこれらの技術を統合し、対話型ロボットエージェント、推薦エージェント、金融市場予測エージェントなどの知的システムとの連携を通し、その有用性を検証する。

【本研究の特色】

既存研究が基本的に特定の応用システム、特定の情報源を前提としているのに対し、本研究における取り組みでは、多様な知的システムと連携できる汎用的かつ持続的な知識収集システムを実現する点に特色がある。また、収集対象とする情報のタイプを事前に特定せず、我々が普段 Web を検索して得ることのできるすべての情報を収集対象とした場合、収集した情報が構造化・半構造化・非構造化データのいずれであるか、また、新たな知識を要求したタスクが何であるかに応じて適切な処理を選択し、収集した情報を知識として構造化する必要があり、そのような知識収集に伴う選択行動の学習もまた本研究課題の実現には欠

かせない要素技術となる。しかしながら、そのような行動学習のフレームワーク自体これまでになく、独創的なものと言える。

【期待できる成果】

本研究課題で実現を目指す技術は、いわゆる汎用人工知能の実現に必要な不可欠なものであり、事前に指定された以上のタスクを遂行可能な知的システムの実現に大きく貢献するものである。また、より具体的な成果としては、より自然に対話の出来る対話型ロボット、想定していない質問にも柔軟に答えることのできる質問応答システム、突発的な事象の発生の下でも適切に金融市場の動向を推定する市場予測システムなどの実現が期待できる。特に、対話型ロボットや質問応答システムなどは、介護や教育などの分野を含め社会的ニーズの高い応用システムであり、それを通じた社会貢献も期待できる。

【研究計画】

研究初年度である 2021 年度においては、以下の技術課題の解決を目指す。

(1) トリガー情報に基づく能動的情報収集基盤のプロトタイプの構築

Web からの知識収集のトリガー情報として、2021 年度はテキスト情報を想定し、能動的情報収集基盤のプロトタイプを構築する。具体的には、既存の分類モデルを用いて新たに与えられた未学習クラスのオブジェクトを分類する問題を対象に、そのクラスに関連するキーワードをトリガー情報として、分類モデルの逐次的な学習に必要な追加情報を Web から能動的に収集する手法を実現する。

(2) 既存概念辞書の動的拡張手法の実現

本申請課題で実現を目指す Web からの持続的な知識収集基盤においては、一度構造化を済ませた知識・概念ベースを、新たに収集した知識を用いて漸進的に拡充していく仕組みが必要となる。2021 年度は、そのうち既存の概念辞書を漸進的に拡充する機構の実現を図る。

(3) 継続学習の基本モデルの構築

本申請課題で実現を目指す Web からの持続的な知識収集基盤が連携を想定するのは、現有知識だけでは遂行不可能なタスクに継続的に取り組む知的システムである。2021 年度は、その具体例として、逐次的に与えられる新たなクラスの画像を分類するというタスクに対応する継続学習システムを想定し、その基本モデルを構築する。

2022 年度以降は、初年度に検討を進めた各課題に対する成果を基礎とし、能動的情報収集のトリガー情報のテキストから画像など他のタイプの情報への拡張、収集した情報のタイプに応じた処理を選択する行動学習システムの実現、収集した情報からの知識構築基盤の拡充などを順次進める。さらに、実現を目指す持続的知識収集基盤と連携する応用システムとして、対話ロボットエージェント、質問応答ソフトウェアエージェント、金融市場予測エージェントなどの知的システムの構築を並行して進め、それらとの連携による実証テストを通して、提案技術の有用性を明らかにする。