

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 18 日現在

機関番号：62615

研究種目：挑戦的萌芽研究

研究期間：2012～2013

課題番号：24650072

研究課題名(和文)物理モデル理解と自然言語処理の統合による試験問題の解答生成

研究課題名(英文) Solving exam questions based on integration of physical model understanding and natural language processing

研究代表者

稲邑 哲也 (Inamura, Tetsunari)

国立情報学研究所・情報学プリンシプル研究系・准教授

研究者番号：20361545

交付決定額(研究期間全体)：(直接経費) 3,000,000円、(間接経費) 900,000円

研究成果の概要(和文)：まず，力学・電磁気・電気回路の各分野のシミュレーションを実施するために，Modelicaと呼ばれるシミュレーションシステム上で試験問題に頻出するバネや滑車などのモジュールを実装した．波動については独自のシミュレーションモジュールを開発した．次に，この拡張された物理シミュレーションモジュールと自然言語処理部を連結するための中間表現である形式表現の記法を検討・考案した．曖昧性解消を実現するために自然言語処理部とシミュレーション処理部間の矛盾検出器を構築した．提案手法を評価するため，実際に実施された大学入試センター模試を対象に答案を作成し，100点満点中39点(偏差値48.3)という結果を得た．

研究成果の概要(英文)：Simulation modules for dynamics, electromagnetics and electrical circuit were developed, for example spring, pulley and so on, in order to realize physical simulation based on the Modelica system. Regarding the wave phenomenon, we developed an original simulator. We also considered and designed "logical form" that connects result of natural language processing and the physics simulator. Disambiguation function was designed based on contrariety detection between the simulation results and given natural language sentences. We evaluated the proposed method with applying to a real exam performed by a cram school; the system got the score 39 among 100 (T-score: 48.3).

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：自然言語処理 物理シミュレーション 曖昧性解消

1. 研究開始当初の背景

従来の自然言語処理研究における文章の理解は、膨大な量の文章コーパスから抽出された統計データによって実現されることが多いため、典型的な質問応答システムは、データベースに記載されている記述的事実に対する解答はできたとしても、実世界の状況や文脈などを参照したり将来に起こる事象を予測して解答することは困難であった。一方で、iPhone に搭載された音声対話システム Siri では、自然言語による機器の操作が可能となっており、IBM は自然言語で出題されるクイズに解答する人工知能を構築している。しかし、いずれも上記の理由により、記述的な知識に関する応答にとどまっている。今後、IT 技術が進歩していくに従い、スマートフォンだけでなく、例えば知能ロボットにおいて、周囲の物理的環境の状況の理解、因果関係に基づく将来予測なども含めた深い対話技術が要求されるようになると予想される。そのため、実世界の状況を画像を介して理解し、物理的な法則に従って変化する事象を物理シミュレーションによってモデル化し、将来を予測する技術の研究が必要となっている。

2. 研究の目的

自然言語による記述的知識の参照だけでは不十分で、物理的環境・状況の理解をおこなう必要のある質問応答システムの応用範囲は事実上無限に存在するが、本研究ではそのための基盤技術の構築と検証を目的として、物理の試験問題に解答するというドメインに焦点を絞る。問題文には物体の幾何学的位置関係を表現する記述があり、かつ、物理的状況の理解から物理的モデルに基づいた将来予測を必要とする事が多い。特に、説明文だけでは解釈に曖昧性があり、将来予測の結果を考慮した上で問題文を理解する必要がある事も多い。そこで、自然言語処理と物理モデル理解の双方が補完しあい、かつ物理シミュレーションに基づく将来予測情報を含んだ質問応答システムに必要な情報処理の仕組みを明確にし、その方法論を確立する。

3. 研究の方法

本研究は以下の4つの研究項目からなる。

(1) 自然言語処理による物理モデル構築と将来予測シミュレーションの確立:

自然言語処理の結果を元に状況の理解を行い、物理シミュレータ内部で物理モデルを再構築する処理を確立する。具体的には Modelica と呼ばれる物理シミュレーション言語を自然言語から生成するべく、中間表現として適切な形式表現を考案する。

(2) 物理シミュレーション上の複数仮説の列挙に基づく解答の選択:

シミュレーションを実行するには、各物体の初速度や初期位置などの初期条件を決定する必要があるが、問題を解くためにはその初期条件が明示されていない事が多い。そこで、認識した状況に応じて初期値を適切に設定する枠組みを設け、複数の初期条件候補を並列に利用して、正答を選択する機能を実装する。

(3) 物理シミュレーションに基づく将来予測を利用した曖昧性解消機能の実装:

自然言語には曖昧性が多く含まれており、明確な記述を指向する試験問題でも、人間の常識を前提とした文章には曖昧性が残る。たとえば「重りを棒に取り付け、持ち上げて離れた」という表記は、「手で持ち上げて、手から離れた」を意味するが、「棒から離れた」という解釈もあり得る。そこで、自然言語理解の複数の仮説のうち、どの仮説を選択すべきか、という処理を(2)と同様に仮説モデルの列挙演算を用いて行う。シミュレーション結果に矛盾がある場合に候補の枝刈りをして正答を探索するシステムの構築を行う。

(4) 模試の実施とその結果に基づくモデル修正

実際に予備校等で実施されている模試をシステムに入力し、回答システムの質を得点という形で可視化する。得点できなかった問題について、システムの改善に求められる要素を検討し、回答モデルの修正を行う。

4. 研究成果

まず平成 24 年度では、物理の試験問題、特に質点の運動に関する力学の問題に解答するためのシミュレーション環境の整備、および自然言語理解に基づくシミュレーションモデルの変換プロセスを開発した。力学の問題解答には、物体の位置、バネやひもの位置・長さ・方向、複数物体間の配置関係などの様々な機械要素の状態を理解する必要があるため、自然言語で記述された構成要素間の関係を、Modelica と呼ばれる機械システム的设计記法に変換し、三次元シミュレータ上での力学シミュレーションを実行可能とする手法を開発した。具体的には物理現象/状態を表現する自然言語を論理的に変換する独自の規則を定義し、回答として答えるべき物理量を指定することで、回答に必要な演算を自動的に生成するアルゴリズムを考案した。また、このシステムを実装するために物理の試験問題に頻出するバネ・滑車・摩擦がある床などの物理モジュールを MultiBody 力学系のモジュールとして実装した。

さらに、自然言語の解釈に曖昧性があり、三次元的な機械要素の配置関係や物理的制約条件等が確定できない状況にしばしば遭遇することになるため、全ての解釈の可能性を列挙し、シミュレーションを実施することで問題文に記載されている内容との矛盾を

検出し、妥当な候補だけを解釈として選択するような曖昧性解消手法を確立した。シミュレーションの結果は物理的状況を説明する全ての状態変数の時系列情報であるため、その中から問題が問うている情報を抽出するために、自然言語表現と数値演算処理の対応関係を定義した。最終的にこれらの要素を結合し、物理の問題を解くための基盤システムが急行に機能することを確認した。

平成 25 年度では、平成 24 年度に構築した力学の問題に解答するためのフレームワークを拡張し、波動・電磁気・電気回路の各分野の問題にも対応可能とするためのシミュレーションモジュールの追加を行った。具体的には、電磁気・電気回路系のモジュールを Modelica のモジュールとして実装し、波動については独自のシミュレーションモジュールを開発した。さらに、この拡張された物理シミュレーションモジュールと自然言語処理部を連結するための中間表現である形式表現の記法を再検討した。平成 24 年度までは、ある単一の物理的状況を記述する表現しか準備されていなかったが、平成 25 年度では「A という状況と B という状況を比較するとどうなるか?」というような質問にも対応するために異なる状況を並列に表現可能とする形式表現記法を交換した。

構築した試験問題解答器の性能を評価するため、実際に実施された大学入試センター模試を対象に答案を作成し、100 点満点中 39 点を得た。これは偏差値に換算すると 48.3 という平均レベルに肉薄するパフォーマンスを得た。

今後の課題として残されているのは、センター試験のように単に選択肢から回答するのではなく、方程式を記述し、理由を述べながら回答するような筆記試験への対応である。現状では、Modelica で記述された三次元物理モデルのシミュレーション時に、複数の拘束条件や物理的関係法則を列挙し、数値演算で将来予測を行った結果のみを参照しているが、数値演算を高速化する際に、Simplification と呼ばれる拘束条件の簡易化を行う処理が機械システムの設計の現場でしばしば用いられることが分っている。この簡易化を行う際に、最も主要な物理的拘束条件を方程式の形で抽出する技術が開発されているため、その技術を応用することで方程式を記述しながら回答をするシステムの構築が可能になるとの議論を行った。具体的には MapleSim と呼ばれるシミュレーションソフトウェアがその機能を持っており、Modelica 記法を受け入れることも可能であるため、今後はこのソフトウェアを中心部としたシステム設計の改善を進めて行く予定である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1 件)

- [1] 稲邑 哲也, 横野 光: "物理モデル理解と自然言語処理の統合による理科問題の回答," 人工知能学会誌, Vol.27, No.5, pp.479-482, 2012. (査読無)

〔学会発表〕(計 6 件)

- [1] Tetsunari Inamura and Hikaru Yokono: "Physics Problem Solver towards Mental Simulation for Real World Oriented Intelligence," proc. of the 4th Annual IEEE International Conference on CYBER Technology in Automation, Control, and Intelligent Systems, pp.591-596, 2014 年 6 月 7 日, 香港.
- [2] 横野 光, 稲邑 哲也: "論理演算と物理シミュレーションの結合による物理問題解答," 2014 年度人工知能学会全国大会予稿集, 2014 年 5 月 13 日, 松山.
- [3] Hikaru Yokono and Tetsunari Inamura: "A Framework of Recognizing Physical Situation in Text Description with Physics Simulation," Proc. of the International Conference on Information Science, Electronics and Electrical Engineering, pp.1090-1094, 2014 年 4 月 28 日 札幌.
- [4] 稲邑 哲也: "大学入試問題を解く「ロボット」と「知能ロボティクス」" 栢森情報科学振興財団第 13 回 K フォーラム(招待講演) 2013 年 8 月 1 日, 高山.
- [5] 横野 光, 稲邑 哲也: "物理シミュレーションと時系列データからの事象認識による物理問題解答," 人工知能学会全国大会予稿集, 2013 年 6 月 5 日, 富山.
- [6] 稲邑 哲也, 横野 光, 新井 紀子: "物理モデル理解と自然言語処理の統合による試験問題の解答生成," 日本ロボット学会学術講演会予稿集, p.RSJ2012AC3M2-1, 2012 年 9 月 19 日, 東京.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

ホームページ等

http://21robot.org/research_activities/science/

6．研究組織

(1)研究代表者

稲邑 哲也 (INAMURA, Tetsunari)

国立情報学研究所・情報学プリンシプル研究系・准教授

研究者番号：20361545

(2)研究分担者

横野 光 (YOKONO, Hikaru)

国立情報学研究所・社会共有知研究センター・特任研究員

研究者番号：60535863

新井 紀子 (ARAI, Noriko)

国立情報学研究所・社会共有知研究センター・教授

研究者番号：40264931

(3)連携研究者

なし