

知識の蓄積による問題解決に向けた 新たな試み

受賞業績 質問応答技術の研究開発とその実用化

武田 浩一 金山 博

日本アイ・ビー・エム(株) 東京基礎研究所

喜安記念業績賞にご選奨いただき大変光栄に存じます。本技術の研究開発チームおよび選定にあたりご尽力いただいた皆様に心から感謝申し上げます。

質問応答は、この10年間で大きな発展を遂げた技術です。米国の情報検索技術の評価型ワークショップにおいて、自然言語で表現された質問に対する解答を生成するタスクとして定義され、日本でも国立情報学研究所が主催する評価型ワークショップ NTCIR を通して研究開発活動が推進されてきました。今回の受賞の対象となったのは、このような技術を利用して、米国の Jeopardy! というテレビの人気クイズ番組に挑戦するために開発された Watson という質問応答システムです。Watson は 2011 年 2 月に全米放映された対戦において、Jeopardy! で歴代最強といえる人間の対戦者 2 名よりも多くの賞金を獲得し、初挑戦で見事に勝利を取ることができました。

Watson の開発は 2006 年に始まり、武田・金山の両名は同プロジェクトが本格化した 2007 年 12 月からコア・チームのメンバとして活動を開始しています。Watson が解決すべき主要な課題は、出題範囲にまったく制限のないオープン・ドメインな質問に対応すること、質問が提示されてから数秒程度で解答を計算できること、解答の確信度を計算して誤答による損失を最小にすること、の 3 点でした。このため各質問に数百程度の解答候補を生成し、それぞれの候補を正解とすべき(情報源から見つかる)根拠の数およびその整合性などをスコア化して確信度を計算するという並列化手法が考案されました。図-1 に 4 つの大きなステップからなる Watson の処理の流れを示しています。Wikipedia やニュース記事をはじめとした大量のテキスト文書を情報源として用いて、質問文と情報源の両方から各種アルゴリズムを用いて抽出した情報をマッチングさせ、正解らしき解答に高い確信度を与える仕組みとなっています。



図-1 質問応答システム Watson の処理の流れ

過去の数万問の正解データを使った機械学習を施すことによって、Watson の性能は飽和することなく着実に向上させることができ、当初は困難だと捉えられていた人間のチャンピオンとの対戦が実現に至ったのです。

米国ではすでに医療分野における保険会社や医療機関において同技術を利用したシステムが実働化しており、現在ビッグデータとして知られる驚異的な情報の蓄積が、質問応答などの技術と融合することで、質問への解答を生成する能力、より広くいえば問題解決能力を実現することを確信できました。国内ではこのような信念のもとで、アカデミアおよび産業界において Watson が実証した質問応答技術の正しい理解とさまざまな応用へのアイデアを喚起できるよう努力してきました。今後はさらに自然言語処理分野や、国立情報学研究所を中心とした「東京大学入試に合格する人工知能分野プロジェクト」に代表される先進的な研究開発活動の推進に貢献できることを願っています。

(2013 年 5 月 17 日受付)

武田 浩一(正会員) takedasu@jp.ibm.com

1983 年京都大学院工学研究科修士課程修了。同年日本アイ・ビー・エム(株)入社。以後、東京基礎研究所において自然言語処理研究に従事。現在同社技術理事。博士(情報学)。本会フェロー。

金山 博 hkana@jp.ibm.com

2000 年東京大学大学院理学系研究科情報科学専攻修士課程修了。同年日本アイ・ビー・エム(株)入社。以後、東京基礎研究所において構文解析、自動翻訳、評判分析などの研究に従事。博士(情報理工学)。