

氏名	Happy Buzaaba		
学位の種類	博士（工学）		
学位記番号	博甲第10247号		
学位授与年月日	令和4年3月25日		
学位授与の要件	学位規則第4条第1項該当		
審査研究科	システム情報工学研究科		
学位論文題目	A Modular and Efficient Approach for Question Answering Over Knowledge Bases (知識ベースに対する質問応答のためのモジュール化された効率的なアプローチ)		
主査	筑波大学 教授	博士（工学）	天笠 俊之
副査	筑波大学 教授	理学博士	北川 博之
副査	筑波大学 教授	博士（工学）	福井 和広
副査	筑波大学 准教授	博士（工学）	乾 孝司
副査	筑波大学 教授	博士（工学）	佐藤 哲司
	(図書館情報メディア専攻)		

論文の要旨

本研究では、知識ベースに対する質問応答に関する研究をおこなっている。近年、ビッグデータやAI技術の進展と普及に伴い、一般的な知識をRDF (Resource Description Framework) などの構造化データあるいはテキストなどの非構造化データなどの形式で蓄積した知識ベースが注目され、さまざまな応用分野において構築、一般への公開が行われている。知識ベースには膨大な情報が蓄積されているため、そこから必要な情報を高速かつ正確に取得することには技術的課題が存在する。とりわけ、検索のために専用言語を用いると、専門的な知識を持たないユーザには利用が難しくなるため、自然言語を利用した知識ベースへの質問応答が期待されている。

この問題に対し、本研究では大きく二つの貢献をおこなっている。一つ目の貢献では、RDF等の構造化知識ベースを対象としたモジュールベースの質問応答手法を提案している。既存の手法の多くは、最新の機械学習モデルをベースとしたエンドツーエンドの手法となっている。これらは高い性能を発揮するが、モデルの学習に多大なコストを要する上、モデルの説明可能性が十分でなく、性能上の問題が発生した際にその原因を特定するのが困難という問題がある。これに対して、本研究では質問応答の過程を1) 実体検出、2) 実体リンク、3) 関係予測の単純なタスクからなる処理パイプラインとして実現するアプローチを提案している。各タスクについて、ニューラルネットワークベースのモデルを含む複数の手法を提案し、各ステップの性能を向上させることで、全体として最新のエンドツーエンドの手法に匹敵する性能を実現している。特に、実体リンクにおいて性能低下の原因が実体の曖昧性に起因することを見出し、それに対して適切な前処理を施すことで、性能が向上することを実験による評価で検証している点は、処理のモジュール化の特性を活かした成果と言える。

二つ目の貢献では、テキストベースの非構造化知識ベースに対するエンドツーエンドの質問応答モデルに対して、埋め込みベクトルの次元削減を適用することで処理を高速化する手法を提案してい

る。エンドツーエンドのモデルでは、質問とそれに対応する解答を分散表現学習モデル等を用いて高次元空間に写像し、質問ベクトルの周辺に存在する解答ベクトルを答えとしているが、モデルの学習に多大なコストを要することが問題である。これに対し本研究では、LSTM オートエンコーダをモデルに組み込み、教師データに対して精度を損なわないような次元削減を行うことで、質問応答の精度を大きく損なうことなく処理の高速化を実現している。

審 査 の 要 旨

【批評】

知識ベースは、汎用あるいは専門的な知識を機械処理可能な形式で蓄積したデータベースであり、近年の AI 技術の進展とともに、幅広い分野で利用が進んでいる。巨大な知識ベースに対してユーザが欲する情報を検索する質問応答の技術は、知識ベースに対する直接的な利用にとどまらず、情報システムのバックエンドにおける処理などに応用が考えられる。特に自然文による質問応答は音声ベースやチャットベースのシステムなどに応用が可能であり、その性能向上が望まれている。

この問題に対し本学位論文では、大きく二つの貢献をしている。一つ目の貢献では、知識グラフとして表現可能な構造化知識ベースに対して、実体検出、実体リンク、関係予測の三つのステップで構成される処理手法を提案している。これにより、最新のエンドツーエンドの手法に匹敵する精度を達成するとともに、性能における問題点の分析や対処のしやすい手法を実現している。

二つ目の貢献では、テキストベースの知識ベースに対するエンドツーエンドの質問応答モデルの処理の効率化に取り組んでいる。従来のモデルの多くは、教師データから質問・解答に対応する数百次元の高次元ベクトルを学習する必要があり、その処理コストが問題となっている。これに対して、提案手法では LSTM オートエンコーダによる次元削減をモデルに組み込み、教師データにおける質問応答の精度になるべく影響を与えないような次元削減モデルを学習する。これにより質問応答の精度を大きく損なうことなく、処理性能の改善を達成している。

以上の成果から、本研究では知識ベースに対する質問応答技術の進展に大きく寄与しており、情報工学上の多大なる貢献が認められる。今後はより複雑な質問応答タスクへの対応や、最新の事前学習言語モデルを取り入れることによるさらなる性能向上などが期待される。

【最終試験の結果】

令和 4 年 2 月 3 日、システム情報工学研究科において、学位論文審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、学位論文審査委員全員によって、合格と判定された。

【結論】

上記の学位論文審査ならびに最終試験の結果に基づき、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。