

(続紙 1)

京都大学	博士 (情報学)	氏名	TAREQ YASER SAMIH ALKHALDI
論文題目	Studies on Question Answering in Open-Book and Closed-Book Settings (オーブンブックおよびクローズドブック設定における質問応答に関する研究)		
(論文内容の要旨)			
<p>Natural language is the medium through which humans mainly communicate and store their knowledge to pass it on. Manual analysis and handling of such massive quantities of natural language is beyond human capabilities, therefore, automatically processing such information through natural language processing (NLP) has been an active field. Searching for information is typically done by asking questions in search engines and going through the returned results to find a specific piece of information in mind, which can be tiresome and time-consuming. A question answering (QA) system can automatically read those documents and return the exact answer immediately, making our lives easier. This thesis focuses on the study of the QA task in different settings.</p> <p>Chapter 1 introduces the background of the QA task and explains in detail the different settings. Considering the complexity of the question, where a question requires one or more reasoning hops, the task is divided into single-hop and multi-hop QA. Considering whether the context is provided to the model alongside the question, the QA task is categorized into open-book and closed-book settings.</p> <p>Chapter 2 presents a review of the recent advancements in NLP that are mainly related to QA. Details are provided on the training procedures for the Transformer architecture, which is the basic block of today's state-of-the-art models. The variations of the Transformer-based models that are used in the thesis are explained, including encoder, decoder, and encoder-decoder models.</p> <p>In Chapter 3, a study for improving multi-hop QA in an open-book setting is conducted. Models for handling multi-hop questions are required to identify and collect important pieces of information from the provided context documents to achieve the answer. Models that utilize the identified supporting sentences have been developed; however, they utilize such sentences ineffectively. They either only focus on the sentences to find the answer and ignore the remaining context or do not use the supporting sentences at all. However, both the sentences and the remaining context may contain important information for the answer. In this chapter, a method is proposed to focus on the supporting sentences more effectively by tagging the sentences while keeping the remaining context and then using the tagged sentences for answer prediction. Additionally, the new method focuses on beneficial signals from context documents by tagging the bridging entities between the documents. Experiments on HotpotQA, a popular dataset for multi-hop QA evaluation, showed an improved answer accuracy of 0.77 EM and 0.98 F1 over models with comparable architecture.</p> <p>In Chapter 4, the setting of closed-book QA was investigated. While the existence and the amount of factual knowledge that models memorize during pre-training have been studied, the mechanism and inner workings are not investigated. This chapter</p>			

investigates how models memorize factual knowledge and their inner workings. Through gradient analysis and corrupting attention components with noise while observing the outcome, the query component in the attention mechanism is found to be the most responsible for finding the answer. Moreover, after pruning parts of the encoder and decoder in turns and monitoring the differences, the encoder is found to be responsible for finding the broad entity embedding, while the decoder finds the exact entity. Furthermore, using different head-scoring techniques, it is demonstrated that certain areas in the model retain more information about factual knowledge than others. Also, it is shown that related entity types are generally clustered in similar areas in the model.

Chapter 5 describes the investigation of multi-hop QA in a closed-book setting and its improvement. The chapter proposes showing reasoning steps that are obtained through explicit and implicit question decomposition to the answering model to improve the performance. The multi-hop QA performance was increased by about 6 EM points on average on MuSiQue, HotpotQA, and 2Wiki datasets. Several possible points for further improving the multi-hop baseline are also identified.

In Chapter 6, this thesis concludes the study of QA models with varying settings and discusses possible trajectories for future research.

(論文審査の結果の要旨)

本論文は、自然言語で与えられる質問に自然言語で回答するタスク、質問応答に様々な設定で取り組み、その研究成果をまとめたものである。本論文では、質問応答を質問の複雑さの観点からシングルホップ質問応答とマルチホップ質問応答に、質問以外のコンテキストの有無という観点からオープンドブック設定とクローズドブック設定に分類し、これらの設定について研究している。主要な成果は以下の通りである。

1. オープンドブック設定におけるマルチホップ質問応答を改善するための研究を行った。マルチホップ質問応答では、回答を得るために提供されたコンテキスト中の重要な情報(サポート文)を特定し利用する必要がある。本研究では、サポート文の情報を有効に活用するにはその文脈も重要であることに着目し、コンテキスト中のサポート文をタグ付けし、予測時に活用する手法を提案した。さらに、サポート文に加えて文書間の橋渡しとなるエンティティをタグ付けすることで、モデルにコンテキスト中の有益な情報に着目させることを提案した。提案手法の有効性を確認するため、マルチホップ質問応答の代表的なデータセットである HotpotQA で実験を行い、同様のアーキテクチャのモデルと比較して回答の精度が完全一致で 0.77 ポイント、部分一致を許す F1 指標で 0.98 ポイント向上することを示した。
2. クローズドブック設定の質問応答において、モデルの内部的な仕組みの分析を行った。言語理解モデルが事前学習時に記憶した事実に知識の範囲はこれまでも研究されてきたが、内部の仕組みに関する研究はなかった。本研究では、モデルが事実に知識をどのように記憶し、また活用するのか調査した。本研究では、勾配分析や注意機構にノイズを加えるといった新しい分析手法を提案した。分析により、注意機構のクエリコンポーネントが答えを見つけるのに最も寄与していることを示した。さらに、エンコーダとデコーダの一部を順番に除外して出力の変化を分析し、デコーダが正確なエンティティを見つけるのに重要であるのに対し、エンコーダは埋め込まれている広範なエンティティを見つける役割を果たしていることを示した。加えて、様々なヘッドスコアリング手法を使用し、モデル内の特定の領域が他の領域よりも多くの事実に知識に関する情報を保持していることを示し、関連するエンティティタイプがモデル内の近い領域にクラスター化されていることを示した。
3. クローズドブック設定におけるマルチホップ質問応答に関する調査を行った。これまでクローズドブック設定におけるマルチホップ質問応答を改善する研究はなかった。本研究では、明示的あるいは暗黙的に質問を分解することで、クローズドブック設定におけるマルチホップ質問への回答を改善した。分析を通じて、マルチホップ質問応答のベースラインをさらに改善するための重要事項を議論した。

よって、本論文は博士(情報学)の学位論文として価値あるものと認める。令和5年8月29日、論文内容とそれに関連した事項について試問を行った結果、合格と認められた。また、本論文のインターネットでの全文公表についても支障がないことを確認した。