

論文内容の要旨

博士論文題目 Identification of Multi-Sentence Question Type and Extraction of Descriptive Answer in Open Domain Question-Answering
(分野を限定しない質問応答における複数文質問の識別と記述的な回答の抽出)

氏名 武智峰樹

(論文内容の要旨)

コンピュータを用いた知的な情報アクセス技術としての質問応答では、インターネットや大規模なデータベースに蓄積された大量の文書からユーザの質問に対する回答を直接抽出する処理が行われる。そこでユーザによって与えられる質問にはジャンルやドメインによる制約が無い。一方産業分野では、電話回線を通じたコールセンターやヘルプデスクなどのサービスや、インターネット上でのオンライン質問応答サービスが数多く存在し、特定の分野の質問だけを扱ったり、人手による回答を行うなど、実用的な手段で質問応答が運用されている。

これら2つのタイプの質問応答は、扱える質問について2つの点で異なる。前者で扱える質問は一文で尋ねることができ、かつ人名、組織名、地名など名詞によって回答できるタイプの質問に限られる。一方、実際の質問応答サービスで扱われている質問には、複数の文を含み、定義、方法、意見など回答として比較的長い記述を必要とするタイプの質問が少なくない。こうした記述的な回答を必要とする質問については、コンピュータによる質問応答ではまだ十分に扱うことができない。本論文が解決しようとする課題はこの問題の解決である。

本論文は、コンピュータによってこのような質問応答が行えるようにするために必要な2つの要素技術、すなわち、ジャンルを問わない複数の文からなるクエリを扱う質問解析と、大量の文書集合から記述的な回答を抽出する回答抽出の技術について論じる。質問解析では、クエリに含まれる質問を抽出し、その質問のタイプを同定する質問タイプ同定について詳しく論じる。本論文は、複数の文からなるクエリに含まれる質問のタイプを適切に同定するための効率的な手法を提案する。一方、回答抽出では、記述的な回答の類型について論じた後、そこから得られた知見に基づき、インターネット上のホームページから手順に関する箇条書きを抽出するタスクについて論じる。また、手順に関する箇条書きを抽出するために必要な箇条書きの特徴についても明らかにする。

氏名	武智峰樹
----	------

(論文審査結果の要旨)

平成19年1月24日に開催した公聴会の結果を参考に平成19年2月20日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

武智峰樹は、本博士論文において、一文による質問、および、単一の名詞による回答を前提とする従来の質問応答システムとは異なり、複数文による質問、および、手順や方法のような複数文による回答を必要とするより広い対象の質問応答を取り上げ、実データの詳細な調査と要素技術の開発を行なった。本論文の貢献は、次のようにまとめることができる。

- 従来の質問応答研究ではあまり取り上げられてこなかった複数文による質問、および、手順や方法のような複数文の回答が必要な質問応答タスクの重要性と有用性を指摘し、一つの実現様式を示したこと。
- 複数文質問のタイプ分類を行い、文レベルのチャンキングを用いて、連続する複数文による質問の識別とタイプの自動分類を行なう手法を提案し、実験による評価を行なったこと。
- 手順に関する質問の回答を獲得するために、Web上の文書から、手順について記述した箇条書きを他の箇条書きと区別して抽出するための方法を提案し、実験により有効性を示したこと。

分野を限定しない複数文による質問と回答を対象とする質問応答タスクに対して、新たな方法論と回答抽出の手法を提案した本研究は、独創性が高く、しかも実用的であり、自然言語処理の応用分野としての質問応答研究において高い貢献があると評価する。

よって、本論文は、博士（工学）の学位論文として価値あるものと認める。