

論文内容の要旨

博士論文題目 A Study on Generic and User-Focused Automatic Summarization
(一般的な要約と視点を考慮した要約に関する研究)

氏名 平尾 努

(論文内容の要旨)

インターネットの急速な発展や大容量の磁気記憶装置の低価格化などによって電子化テキストが氾濫している。情報の洪水といわれる状況下では、我々が本当に欲している情報を手にいれることが困難である。こうした状況を背景として、情報の取捨選択を容易にするための研究が行われている。例えば、「情報検索」、「情報抽出」、「質問応答」などがそうである。近年、こうした研究の中でも特に「自動要約」に対する関心が高まってきている。文書が要約されることによって、概要把握にかかる時間は削減でき、欲する情報か否かの判断も迅速にできるようになる。

本研究では、自動要約に関する以下の3つのトピックを対象とする。

(1) 多数の素性を用いた高性能な単一文書要約手法 (2)(1) の拡張としての高性能な複数文書要約手法 (3) 質問応答システムの解答の根拠を示すための要約手法

(1)に関しては、Support Vector Machine を用いた単一文書要約手法について述べる。従来より、文書要約では多数の素性を統合することが効果的であるといわれている。しかし、素性の数が多い場合には、スコア関数手法では人手によりパラメータの重みの最適値を決定することが困難である。また、機械学習手法の一種である決定木学習も多数の素性を用いた場合には成績が悪いことが知られている。そこで、多数の素性を用いても過学習しにくいとされる Support Vector Machine を用いた。さらに、学習した素性の重みを解析することによって、要約のために有効な素性を明らかにした。

(2)に関しては、(1)の拡張として Support Vector Machine を用いた複数文書要約手法について述べる。複数文書要約も単一文書要約と同様に多数の素性を統合することが効果的である。さらに、文書集合に特徴的な素性を導入する必要がある。素性の数は単一文書の場合よりも多くなる。よって、単一文書要約において多数の素性を効果的に扱えた Support Vector Machine を用いた。また、複数の文書から作成した要約には、冗長な表現が多いということが言われている。そこで、冗長性を削減する手法としてよく用いられる Maximum Marginal Relevance を用いて冗長性削減の有効性を調べた。

(3)に関しては、質問応答システムの解答の根拠を提示する Question-Biased Text Summarization について述べる。質問応答システムは質問に対して解答となる文字列(単語や句など)を返すシステムである。いわゆる文書検索に代る技術として注目されている。しかし、システムの出力が常に正しいとは限らないので利用者はシステムの解答の正誤判断を行わなければならない。そこで、解答候補となる単語と質問に含まれる単語に着目し、それらが近接するパッセージを抽出することで解答の正誤判断に必要な文脈を保持した要約を作成する。

氏名	平尾 努
----	------

(論文審査結果の要旨)

平成14年7月24日に開催した公聴会の結果を参考に平成14年8月21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。

平尾 努は、本博士論文において、大きく3つの視点から文書要約について研究を行った。すなわち、単独の文書に関する要約、同一の話題を扱った複数文書に関する要約、および、質問応答システムの解答としてその根拠を端的に示すための応用に特化した要約である。主として機械学習による手法を利用して、従来手法を上回る要約精度を達成するとともに、次のような事項を明かにした。

1. Support Vector Machine を用いることにより、従来よりも多数の素性を統合することにより、要約精度を向上させた。一方、素性の数が多い場合には、人手によりパラメータの重みの最適値を決定することが困難であるため、SVMによって学習した素性の重みを解析することにより、要約のために有効な素性を明らかにした。
2. 複数文書要約も単一文書要約と同様に多数の素性を統合することが効果的であるが、さらに、文書集合に特徴的な素性を導入する必要がある。よって、単一文書要約において多数の素性を効果的に扱えた Support Vector Machine を用い、高い要約精度を達成した。また、複数の文書から作成した要約に多く含まれると言われる冗長性を削減する手法として Maximum Marginal Relevance を用いて冗長性削減の有効性を調査した。
3. 質問応答システムの解答の根拠を提示する Question-Biased Text Summarization について研究した。質問応答システムが質問に対して回答する出力は常に正しいとは限らない。利用者がシステムの解答の正誤判断を容易に行わなうことができるためには、解答候補となるパッセージ自体の要約を作成すれば、利用者の助けになる。このように目的に特化した要約を機械学習を用いて実装し、被験者による評価実験を行い、有効性を明かにした。

以上のように種々の利用形態を対象に要約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、文書要約の分野において高い貢献があると評価する。