

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Enhanced models for query-oriented extractive summarization
著者(和文)	森田一
Author(English)	Hajime Morita
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9659号, 授与年月日:2014年9月25日, 学位の種別:課程博士, 審査員:奥村 学,新田 克己,山田 誠二,本村 陽一,高村 大也
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9659号, Conferred date:2014/9/25, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

論文審査の要旨及び審査員

(2000字程度)

報告番号	乙 第 号	学位申請者	森田一	
	氏 名	職 名	氏 名	職 名
論文審査員	主査 奥村学	教授	高村大也	准教授
	新田克己	教授		
	山田誠二	連携教授		
	本村陽一	連携准教授		

Web の進歩により、Web 上に存在する文書は既に個人が読むことができる量をはるかに超え、人々が手にすることができる情報は今もなお飛躍的に増加し続けている。文書要約は必要な情報を保ちつつ、読まなければならない文書量を減らすことを目的としており、文書要約技術は情報過多に対抗する手段の一つとして期待されている。中でも、クエリ指向複数文書要約は、ユーザの指定したクエリに関連した情報を大量の文書集合から集めて提示することを目的としている。クエリ指向複数文書要約技術を利用することで、ユーザは大量の文書を読む代わりに生成された要約を読むことで必要な情報を効率よく得ることができる。

本論文は、クエリ指向の抽出的複数文書要約のための新しい手法の開発を目的としており、「Enhanced Models for Query-Oriented Extractive Summarization (クエリ指向の抽出的要約モデルの高度化に関する研究)」と題し、英文全5章より構成されている。

第1章「Introduction (序論)」ではまず、本研究の背景を述べ、従来の文書要約の問題点について触れている。クエリ指向要約のモデルは、文書中の情報とクエリとの関連の強さを計算するモデルと、関連の強さに基づき要約を生成するモデルからなるとし、本研究の目的が、その両モデルを改善することによりクエリ指向要約を高度化することであると述べている。

第2章「Text summarization (テキスト要約)」では、従来の文書要約技術および、関連する分野の従来の研究について述べている。

第3章「Query snowball: a co-occurrence-based approach to multi-document summarization for question answering (質問応答のための共起に基づく複数文書要約)」ではまず、クエリの表現方法の拡張として、要約元文書集合中で単語の共起に基づきクエリ関連度を計算する手法である Query snowball を提案している。この手法では、クエリを文書中の各単語に対するクエリ関連度として表現することで、ユーザから与えられたクエリ中の単語が含まれない文に対してもクエリとの関連度を計算することが可能であることが利点であると述べている。提案手法では、最初に要約元文書集合中での単語の共起関係をグラフとして表現し、次に、クエリ単語との直接的な共起および、他の単語を経由した間接的な共起を元に、クエリ単語と直接共起した単語、間接的に共起した単語の順にクエリからスコアを伝播させていくことによりクエリ関連度を計算している。次に、文のクエリ関連度を計算するための手法として、単語の組み合わせに対してクエリとの関連度を計算する手法を提案している。単語の組み合わせを、スコアおよび冗長性を計算する単位として利用することで、クエリに関連する複数の話題を要約に含めやすくできるという利点があると述べている。評価実験の結果、比較手法と比べて高い質の要約が生成できたことを示している。

第4章「Subtree extractive summarization via submodular maximization (劣モジュラ最大化アルゴリズムを用いた文抽出と文圧縮に基づくクエリ指向要約)」では、ユーザの与えるクエリに対して、文の選択と文の圧縮を同時に行い、必要な個所のみを要約として提示するための部分木抽出要約モデルを提案している。文の選択と文の圧縮を同時に行うことで、生成する要約から冗長な個所とユーザのクエリに関連しない個所の両方を除くことができる点が提案手法の利点であると述べている。次に、提案した要約モデルを劣モジュラ最大化問題として定式化し、最大化問題を近似的に解くための貪欲アルゴリズムの提案と貪欲アルゴリズムの性能についての理論的な解析を行い、提案した近似アルゴリズムが最適解に対して一定の近似率を保証することを示している。劣モジュラ最大化問題としての定式化および提案したアルゴリズムにより、提案モデルは、ユーザのクエリに対して応答する上で十分高速に動作することが可能となると述べており、評価実験の結果、比較手法と比べ性能が向上していることを示している。

第5章「Conclusion and future work (結論)」では、本研究で達成した内容について述べ、今後の課題について述べている。

以上を要するに、本論文は、クエリ指向の抽出的複数文書要約モデルを高度化するための手法を提案し、評価実験を通じてその有用性を確認しており、工学上貢献するところが大きい。よって、博士(工学)の学位を授与するに十分な価値をもつものと認められる。