

(続紙 1)

京都大学	博士 (情報学)	氏名	原島 純
論文題目	Studies on Re-ranking and Summarizing Search Results (検索結果の並べ替えと要約に関する研究)		
(論文内容の要旨)			
<p>本論文は、検索エンジンが返す検索結果から情報を効率的に取得するための手法・システムについて論じたものである。具体的には、検索結果を並べ替える手法、検索結果を要約するシステム、日本語文を圧縮する手法を提案している。論文は、これらの三つの手法・システムについて各々一章を割り当て、序論と結論の章を合わせて全五章から構成されている。</p> <p>第一章は序論である。現状の検索エンジンが抱える問題、この問題に関連する研究トピックである情報検索と文書自動要約、本論文の貢献について説明している。</p> <p>第二章では、検索結果を並べ替える手法を提案している。導入部では、ユーザの情報要求に見合う検索結果を得るのが難しいこと、この問題の解決を図る適合性フィードバックと呼ばれる技術があること、既存の適合性フィードバックの手法がテキストに表層的に出現する単語の情報のみを用いて並べ替えを行っていることを説明している。次に、提案手法の基礎をなす技術として、言語モデルに基づく検索結果のランキング手法と、トピックモデルについて解説している。その後、これらの技術に基づいて、テキストに表層的に出現する単語の情報だけでなく、テキストに潜在的に現れうる単語の情報も用いて並べ替えを行う手法を提案している。実験の節では、評価型ワークショップのテストセットを用いた実験とその結果について報告している。実験の結果、提案手法が、表層情報のみを用いる既存手法と比べて、初期検索結果をより適切に並べ替えることが示されている。</p> <p>第三章では、検索結果を要約するシステムを提案している。導入部では、検索結果を要約するためには、一般的な複数文書要約と違い、要約対象となる文書集合に含まれる複数のトピックを考慮する必要があることを説明している。また、各トピックについて要約を生成するためには、要約内の冗長性と要約間の冗長性という二種類の冗長性を削減する必要があることを説明している。次に、トピックモデルを用いて、検索結果に含まれるトピックの検出、各トピックへの文書の分類、各トピックに対する要約の生成を行うシステムを提案している。提案システムでは、トピックモデルを用いて推定した各トピックに対する各文書や各単語の重要度を用いることで、トピック毎にそのトピックに特徴的な要約を生成している。実験の節では、被験者実験とその結果について報告している。提案システムを被験者に実際に利用させた結果、システムによって生成された要約が、クエリに関する情報を収集する上で有効であったことが示されている。</p> <p>第四章では、日本語文を圧縮する手法を提案している。導入部では、コンパクトな要約を生成するためには、重要文を抽出するだけでなく、文を圧縮する必要があることを説明している。また、日本語文が単語列とも文節列ともみなせるため、既存の圧縮手法が単語ベースの手法と文節ベースの手法に分けられることを説明している。次に、単語ベースの手法と文節ベースの手法について解説し、単語ベースの手法では文法的な圧縮文を生成するのが難しいこと、文節ベースの手法では文節内の不要な単語を除去できないことを指摘している。その後、文節を構成するための制約を緩和する</p>			

ことでこれらの問題を克服する新しい圧縮手法を提案している。実験の節では、新聞記事コーパスを用いた実験とその結果について報告している。実験の結果、提案手法が、単語ベースの手法や文節ベースの手法より原文の情報を保持した圧縮文を生成できることが示されている。また、提案手法が、単語ベースの手法より文法的な圧縮文を生成できることが示されている。

第五章は結論である。本論文のまとめと今後の課題について述べている。

(論文審査の結果の要旨)

本論文は、検索結果からの情報収集を支援する手法・システムの研究成果をまとめたものである。得られた主要な成果は以下の通りである。

1. 従来の適合性フィードバックの手法は、テキストの表層情報のみを用いて検索結果の並べ替えを行っていた。一方、本論文では、テキストの表層情報だけでなく、潜在情報も用いて並べ替えを行う手法が提案されている。テキストの潜在情報は、トピックモデルを用いることで、これを取得している。提案手法では、表層情報と潜在情報の両方を用いてテキスト間の類似度を測り、検索結果を並べ替えている。本論文では、大規模なテストセットを用いた実験とその結果について述べられており、提案手法が従来手法より優れた手法であることが示されている。

2. 既存の検索エンジンは、検索された文書のリストを提示するのみにとどまっている。これに対し、本論文では、検索結果と共に検索結果の要約を提示するシステムを提案している。提案システムでは、検索結果に含まれるトピックが自動で検出され、各トピックについて要約が生成・提示される。これにより、ユーザは、提示された要約に目を通すだけでクエリに関する情報を収集することができるようになる。本論文では、ユーザに提案システムを実際に利用させ、その有効性を調査している。また、調査の結果、提案システムがユーザの情報収集に有効であることが確認されている。

3. よりコンパクトな要約を実現するためには、要約中の各文を圧縮する必要がある。しかし、日本語文に対する従来の圧縮手法は、文節の扱いに関して種々の問題を抱えていた。例えば、単語ベースの手法は、文節をまったく考慮しないため、文法的な圧縮文を生成するのが難しいという問題があった。また、文節ベースの手法は、文節を一つのユニットとして扱うため、文節内の不要な単語を除去できないという問題があった。本論文では、これらの問題を解決するため、文節を構成するための制約を緩和する手法を提案している。すなわち、提案手法では、文節内の不要な単語を落としつつ、原文の文節をできるだけ保持できる。これにより、従来手法では成し得なかった、文節の圧縮と、圧縮文の文法性の保持の両立を実現している。また、圧縮文の情報量や文法性についての実験を通して、提案手法が従来手法より優れた手法であることが確認されている。

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。また、平成25年2月21日実施した論文内容とそれに関連した試問の結果合格と認めた。