審 査 の 結 果 の 要 旨

氏　　名　　正 田 備 也

本論文は「リンク情報を利用したWeb文書クラスタリングに関する研究」と題し、大量のWebページを任意の粒度のクラスタに分割する新たな手法とその文書検索への応用やネットサーフィンのナビゲーションへの応用により、その有効性を論じたものであり、6章から構成されている。

第１章は「総論」であり、本研究の取り扱う問題および研究の目的と論文の構成について述べている。研究対象とするWeb文書とハイパーリンクの特質をまとめ、論文の目的として、Webページをクラスタリングする新たな手法の提案と、提案手法をネットサーフィンのナビゲーションおよびWeb検索へと応用する方法の提示に設定している。

第２章は「研究の背景」と題し、Webにおける文書の相互参照の偏在というグラフ構造を、他のリンク構造を有する情報と比較して論じている。そしてリンク解析および文書クラスタリングに関する先行研究を紹介した上で、新たに提案しようとするクラスタリング手法が、全体としては強連結でないものの相互参照が至る所に存在するというWebグラフ本来の姿を可能な限り生かしている、という特性を持つことを述べている。また、クラスタの個数ではなくクラスタの大きさを意識した粒度制御の機構を持ち、同時にスケーラビリティの確保も考慮して構想された手法という点で、新規性を持つことを示している。

第３章は「パラメータ化された連結性に基づくクラスタリング」と題し、提案手法を詳細に説明している。この手法では、まず有向グラフとしてのWebに新しい距離の概念を導入する。次に、その長さが閾値パラメータと呼ばれるパラメータ値以下の相互参照でつながったWebページのみを同じクラスタにまとめる。相互参照の長さは、その経路上のリンク重みの総和で定義する。この定義より、クラスタの直径は必ず閾値パラメータの２倍以下となる。つまり、得られるクラスタに一定の定量的性質が保証されるという特徴が利点となっている。

第４章は「ネットサーフィンのナビゲーション」と題し、効果的なWeb文書間のリンクの巡航を目指す新奇探索型サーフィンという新しい概念を提案している。Web空間のリンク構造は次々と新しいWebページを見続けにくい構造を持つことを示し、これを改善するためのネットサーフィンの新たなモデルを提案している。このモデルは、PageRankのモデルより現実的でありながら、一定の数理的解析を許容する点で優れた特徴を持つ。そして、提案手法の与えるクラスタリングの結果を利用したナビゲーションをこのモデルに組み込み、モンテカルロ法によるシミュレーション実験を行ってその効果を示している。

第５章は「Web検索」と題し、提案手法のWeb検索への応用を論じている。検索への適用においては、個々のクラスタを、Webページが仮想的にテキスト情報を交換しあう圏域と考える。つまり、多くの適合ページを含むクラスタでは、メンバーであるWebページの適合度が強調されるよう、各ページから抽出された特徴量としての文書ベクトルを

修正する。そして、提案手法によって得られたクラスタリングの結果が、検索結果順位の的確な付け直しに寄与するか否かを調べた。国立情報学研究所の提供するNTCIRテストコレクションを用いて実験を行い、その結果として検索性能の向上を発揮できたことを示している。

第６章は「結論」であり、本論文の成果をまとめている。まず、新しいクラスタリング手法を提案して、提案手法の与えるクラスタリングの有効性が、ネットサーフィンのナビゲーションに寄与することが確認できたことが第一である。さらに、Web検索については、検索性能を向上させる方向に動かすことができることを示した。また今後の発展的課題として、前者についてはクローリングへの応用、後者については、適合度フィードバックや自動質問拡張など、新しい情報検索手法の内部にクラスタリング結果を直接導入する方法の提案が考えられることを述べている。

以上のように、本論文は、Web文書のグラフ構造に基づいて、所望の粒度のWeb文書クラスタを生成するための新しい手法として「パラメータ化された連結性」に基づく一般性を持ったクラスタリング手法を提案した研究で、Web文書クラスタの利用により効果的なネットサーフィンを行うことができることをシミュレーションで示すとともに、テストコレクションを用いて情報検索の性能の向上を示すことによって、提案の有効性を実証した研究であり、電子情報学に貢献するところが少なくない。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。