

審査の結果の要旨

氏名 古川 忠延

本論文は「ウェブにおけるユーザの行動に着目した嗜好の分析と推薦手法の研究」と題し、6章から成る。

第1章は「序論」であり、ウェブは近年、情報流通、共有の重要なインフラになってきたが、情報過多、氾濫の状況も見られるようになり、利用者に必要な情報を適切に提示する推薦システムが必要になってきているという、本研究の背景を記している。そして本論文は、このような情報推薦機能を実現する上で有用な情報を、ユーザ行動の蓄積から抽出することを目的とした、以下の3つの研究について記している。(1)ブログにおける社会ネットワークと閲覧行動の分析、(2)ブログ上の情報伝播に基づく話題抽出、(3)ソーシャルブックマークを対象としたユーザ行動の遷移に基づく推薦手法の研究。

第2章「ブログネットワークの調査」では、次章のブログ閲覧行為を判別する際に用いる素性である、ブログのコメント、トラックバックといった繋がりが有する性質を、ネットワーク分析の視点から調査している。文献調査と共に実ブログデータによる解析も行っており、関係を持つネットワークの Small World 性、Scale Free 性、推移律、双方向性、QAP(Quadratic Assignment Procedure)相関性、その他性質について示している。そして、ブログ上では多くの繋がりを集めるハブ的なユーザが存在すること、関係の種類によっては繋がりが生じると逆方向の繋がりが形成される可能性が高いこと等を示している。

第3章「ブログにおける閲覧行動分析」では、実ブログデータ解析により定期的な閲覧関係(RR 関係)を中心にして、その価値や社会的関係との関連性を調べている。そして、例えばユーザは3度以上訪問したことのあるブログのうち50%に対して、システムへのログイン5回に1度以上の頻度で訪れ、同80%のブログに対してはログイン13回に1度以上の頻度で訪問しているといった統計量を見出している。次いで、RR 関係がどのように情報の普及に影響しているのか、どのような情報が RR 関係により伝播しやすいかを調べている。その結果として、幾つかの情報は RR 関係を通じて伝播しやすいことを示し、概して RR 関係上では RR 関係がない場合と比較して、情報が短い期間で、かつ高確率で伝播しやすいことを示している。上記のような RR 関係は種々の素性を基にして、決定木を用いる機械学習によりある程度判定可能であることを示している。

第4章「話題の伝播モデル分析」では、ブログ間の話題伝播が、語の力とブロガーの力によって説明できることを前提として、伝播の情報から議論の連なりやすい語を重要語として判別する手法を提案している。これは世間一般で良く知られた話題と比較し、人づてに伝播しやすい話題を判別する手法となっている。語の力とブロガーの力を分離して取り扱うために、特異値分解を用いる方法を導入している。実ブログデータを用いた実験により、瞬発性や継続性を持つ語を重要語として定義することで、規模に依らず話題性のある語を重要語として判別できることを示し、これは出現頻度の変化だけでは判別しづらい語にも対応出来ている。

第5章「行動の遷移モデルに基づいた嗜好の予測」では、ソーシャルブックマークのデータ

を主な対象にして、ウェブページに対するユーザ間の登録順序、並びにユーザが登録するウェブページ間の順序に基づいて、将来ユーザが登録するであろうウェブページを予測する手法を提案している。これは順序関係を連続時間マルコフ連鎖によってモデル化することで、あるユーザから別のユーザへの登録が伝播する可能性を推移確率によりスコア付けする手法となっている。ユーザ間遷移を考慮する手法では、先駆的イノベータが存在するソーシャルブックマークのデータにおいて高い予測精度を示し、更にウェブページ間の遷移とウェブページのクラスタリングを加えることで、よりユーザ毎の嗜好の遷移に従った予測ができることを示している。

第6章は「結論」であり、本論文の成果をまとめている。

以上のように、本論文は情報過多、氾濫の問題が現れてきているウェブ情報空間において、利用者に必要と推測される情報を推薦する機能が必要となってきたという背景の下で研究を行い、そのような新しい情報推薦に必要な情報を、ユーザ行動の蓄積から抽出する以下の研究結果を提示している。第一はブログにおける閲覧行動を分析し、定期的閲覧関係（RR 関係）をブログに存在する素性を基にして機械学習により判定する方法を示し、RR 関係を通じて情報が伝播しやすいことを明らかにしている。第二に、ブログ間の話題伝播において、話題語の力とブロガーの力を行列の特異値分解によって分離して取り扱えるようにし、語の出現頻度の変化だけでは判別しにくい瞬発性や継続性を有して伝播する語と重要話題語を判別する手法を示している。第三に、ソーシャルブックマークを主な対象にして、ウェブページに対するユーザ間の登録順序、並びにユーザが登録するウェブページ間の順序を連続時間マルコフ連鎖によってモデル化し、あるユーザから別のユーザへの登録が伝播する可能性を推移確率によりスコア付けする手法を示している。これらの解析や手法は実データを用いて効果を実験的に示しており、ウェブ情報推薦の新しい可能性を提示したものとして、この分野に少なくない貢献を果している。すなわち、本研究は情報理工学に関する研究的意義と共に、情報理工学における創造的実践に関し価値が認められる。

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。