学位論文題名

ビッグデータとしての統合行動履歴に基づく
興味推定法に関する研究(A Study on Inference of User Preferences based on Integrated Behavioral
Logs as Bigdata)

学位論文内容の要旨

インターネットの普及とあいまって情報爆発が叫ばれて久しく、インターネットから利用できるサービスやコンテンツは現在もなお増大を続けており、それら大量のコンテンツからのユーザの選択や発見を支援する技術の重要性が増している。このため、ユーザの行動履歴を分析することで、その興味を推定し、選択や発見を支援する研究が活発に行われている。従来の研究では、複数のユーザの行動履歴の類似性に基づく協調フィルタや、ユーザが閲覧等により消費したコンテンツ内容の傾向分析に基づくコンテンツベースフィルタ、及びそれらを合わせ用いたハイブリッドタイプ等の研究が知られている。しかし、これらの研究において、各ユーザ及び各コンテンツに対する一定量以上の履歴が蓄積されなければ推薦精度が向上しないという、Cold-start(First-rater)という問題が知られている。サービス提供者の観点から本問題を捉えると、すなわち、サービス開始初期や、新規追加コンテンツに対して、推薦精度を直ちに向上できないことを意味するため、解決が強く求められる。Cold-start問題に対しては、コンテンツベースフィルタが比較的有効であることが知られており、ベイジ確率や機械学習等の導入などを踏まえ、より正確にユーザの興味を推定する取り組みが進められているものの、いわゆる教師データとなるべき履歴の量が少ない場合においては、推薦精度の向上には自ずと限界があることから、さらなる改善が求められる。

一方で、近年急速に普及するスマートフォンに代表される、インターネットに接続する機器の高性能化、多様化、センサ種類の増加から、様々な種類の行動履歴データを常時且つ同時収集することが容易となっている。このような行動履歴データは、量 (Volume)、生成速度 (Velocity)、多様性 (Variety) の観点から、従来の技術では処理困難とされ、近年活発な研究対象となりつつある「ビッグデータ」としての特性をとりうる。ビッグデータの解析方策としては、分析処理アルゴリズムを分散処理可能とすること、入力データの逐次処理化(オンライン学習化)などの取り組みが知られている。

本研究では、多様な機器・サービスから常時収集された行動履歴データを統合行動履歴と呼び、これを分析対象とするとともに、ビッグデータとみなしうる行動履歴データの量や種類を増した上で統合的に解析を行うことで、Cold-start問題への対処とし、コンテンツの新規追加や変化が多いサービスにおいても高い推薦精度を維持可能とするようなユーザ興味推定技術の実現に取り組む。統合行動履歴の分析における課題として、多種行動履歴データの分析、常時収集により含まれるノイズへの対処、および逐次追加される大量データの処理を挙げることができる。本研究では、これらの課題に対して、コンテンツベースフィルタで広く利用される、ベクトル空間モデルを用いた手法として、複数の統計的仮説検定を統合して用いる手法を応用した「偶然からの乖離確率による分析」を提案する。提案法について、適用先として想定されるサービスを前提とした実験条件を設定し、結果の評価を通じて提案法の有効性を確認した。また、実験環境を提供するためのアプリケーションを開発した。

本論文は7つの章から構成される。第1章では、本研究の背景、目的について説明している。

第2章では、本研究に関する従来研究について説明している。まず、行動履歴の分析によるコンテンツ推薦の従来研究について概観し、代表的な課題のひとつとされている Cold-start 問題について詳述している。次に、近年のスマートフォンに代表されるインターネットに接続する機器の高度化や搭載センサの増加から収集が容易となった多様な行動履歴について概説するとともに、近年活発な研究が行われているビッグデータについても説明を加え、常時収集された多様な行動履歴データがビッグデータとなりうることを述べる。また、ビッグデータに対する分析処理の従来研究についても説明している。

第3章では、多種類の行動履歴を、一律に偶然からの乖離と捉え、統一的に扱うという、本論文における提案手法について説明している。まず、提案手法のアルゴリズムおよび理論について述べた後、統合行動履歴について改めて詳述し、Cold-start 問題に対する有効性について考察する。さらに、提案手法において想定される問題点と対処について説明している。

第4章では、提案手法をニュース推薦に適用した実験と評価結果を示す。ニュース推薦は、新規追加コンテンツが推薦対象として重要で Cold-start 問題の影響を強く受けることを実験の背景として述べる。次に、1日1度のニュースの閲覧を仮想的に行った被験者による実験とその評価から、提案法が Cold-start 問題に対して有効であることの評価結果を説明している。

第5章では、提案法が対象とする、行動履歴の種類を増やし、GPS 移動履歴からユーザの興味推定を行った実験の結果を示す。約1か月間収集した複数の被験者の GPS 移動履歴から、グルメに関する興味を推定する実験において、提案法が従来法に比べて高い推定精度を示したことを検証結果として述べている。

第6章では、提案法を飲食店推薦にも適用し、GPS 移動履歴と、情報閲覧履歴の統合分析を行った実験と評価について論じている。実験の背景として、飲食店は、開廃業率が高い上に、メニュー等のサービス内容の変化も大きいため、Cold-start 問題の影響を強く受けることを述べた上で、実験結果の評価を通じて、提案法が Cold-start 問題の克服に大きく寄与しうることを述べている。さらに、統合行動履歴において、ビッグデータの特性として知られる、量 (Volume) 及び生成速度 (Velocity) への対処法として、入力データ毎のオンライン学習処理時間を評価し、大規模ユーザ利用時にも、提案手法が有効に機能しうることを説明している。

第7章では、まとめと今後の課題について述べている。

学位論文審査の要旨

主査	教授	水田	正弘
副査	教授	赤間	清
副査	教授	高井	昌彰
副査	准教授	南	弘征

学位論文題名

ビッグデータとしての統合行動履歴に基づく 興味推定法に関する研究

(A Study on Inference of User Preferences based on Integrated Behavioral
Logs as Bigdata)

本論文は、スマートフォンなどの携帯機器やモバイルサービスなどを利用するユーザについて、サービス提供側で得られる使用履歴や、携帯機器内の機能を介して得られる行動履歴などの各種履歴情報を統合的に解析し、ユーザの興味に対する推定へ応用し、サービス精度の向上を目指した研究について論じている。

巷間、ビッグデータと呼ばれるような大量データの解析や分類において、機械学習の分野では数多くの研究成果が報告されているが、その多くは、既存履歴を教師データとしているため、履歴数が相対量に満たない場合には、十分な成果を得にくいという Cold-start 問題を有している。そこで本論文では、さまざまな機器・サービスからそれぞれ得られるユーザ履歴を一括して扱うこととし、教師データを増加させることで Cold-start 問題の克服を目指すとともに、機械学習における問題点として知られている、変化への追従性についても評価検討を行っている。具体的には、入力となる各種履歴データをベクトル空間上にあらわした上で、複数の統計的仮説検定手法を統合的に適用する「偶然からの乖離確率による分析」の適用を提案し、想定サービスを模した数値実験による評価と、実験環境を具現化するためのアプリケーションの開発を行っている。

本論文は7つの章から構成されており、第1章では、本研究の背景、目的について説明している。第2章では、本研究の基盤技術あるいは関連技術となる従来研究について述べている。具体的には、行動履歴を入力としたコンテンツ推薦技法に関する従来研究ならびに Cold-start 問題を概観したのち、携帯機器の高度化や搭載センサの増加が多様な行動履歴の入手可能性をもたらしたことや、関連研究として昨今注目度の高いビッグデータ処理の概要と本研究で用いる履歴データとの関連性にも言及している、

第3章では、多種類の行動履歴について、いずれも一律に、偶然からの乖離と位置付けることで統一的に扱うという、本論の根幹を成す着想ならびに提案手法を子細に述べている。実際には、提案手法の理論的背景と実装アルゴリズムの説明ののち、対象となる履歴データならびに Cold-start 問題に対する有効性について考察を加えている。また、提案手法で想定される問題点と対処についても説明を加えている。

第4章では、Cold-start 問題の影響を受けやすいとされるニュース推薦に対して、本研究に基づく提案手法を適用した実験と評価結果に言及し、その有効性について検討を行っている。さらに第5章では、分析対象とする履歴の種類を増やし、GPS 移動履歴からユーザの潜在的な興味を推定を試

みた実験の結果について述べ、その妥当性を考察している。第6章では、著者の本務先において想定されているサービスの一つである、飲食店の推薦に本提案手法を適用し、GPS 移動履歴と情報閲覧履歴を統合して分析を行った実証実験とその評価から、複数異種の履歴を用いることによって、より適切な推薦結果が実時間で得られることを実際に示し、本研究の有効性を検証している。

第7章では、本論文のまとめと、なお残されている課題について述べている

これを要するに、著者は、既知の情報推薦技術における問題点に対して、複数異種の履歴情報の統合的な解析技法を提案するとともに、その有効性を実証しており、情報科学に対する貢献は大なるものがある。よって、博士(情報科学)を授与するに値すると認める。