

( 続紙 1 )

京都大学	博士 (情報学)	氏名	田頭 幸浩
論文題目	Practical Web-scale Recommender Systems (実用的なWebスケール推薦システム)		
(論文内容の要旨)			
<p>本論文は、ウェブサービスの各ユーザに向けて個別化されたコンテンツを提示する大規模な推薦システムを、高速・高精度に実現するにあたって直面する課題を解決するための方法について研究した結果をまとめたものであり、6章から構成されている。</p> <p>第1章は序論であり、本研究の目的とその内容を概観している。ウェブサービスにおける推薦システムの普及を背景として、主に大規模な広告配信システムを実用レベルで実現する際に直面する2つの重要な問題、すなわち、推薦対象が時々刻々と変化することから膨大な候補の中から少数のものを短時間で選択する問題、ユーザの行動履歴などさまざまな情報からユーザの興味を適切にコンパクトに表現する問題を挙げている。</p> <p>第2章では、関連研究についてまとめている。推薦システムにおける代表的な問題設定である評価値予測問題の典型的な定式化である行列分解とその解法を紹介した後、推薦システムの代表的な評価指標や、ペアワイズ・リストワイズといったモデル化のアプローチを概観している。また、推薦システムにおいて特に興味があるトップ数件の対象の検索問題と、対象の表現法として高次元の疎表現・低次元の密表現等について関連研究を整理している。</p> <p>第3章では、一千万件近くの極めて多数の広告候補の中から、数百程度の適切な候補を効率的に取り出すための方法として、転置インデックスと機械学習を組み合わせた方法を提案している。広告やウェブページと、これらと対になる検索クエリは、含まれる単語によって表現されることが多い。通常これらは高次元で疎な表現となるため、転置インデックスとは相性がよいものの、語彙の異なりから類似度を計算することが困難であった。そこで、本研究では広告などに含まれる単語と、クエリに含まれる単語を対応づける変換表をデータから学習する問題を定式化することを提案している。実際の広告配信システムに提案手法を適用することにより、既存手法と比較して広告クリック率や収益を上げることができることを確認している。また、学習された変換表からは意味的にも妥当な対応が得られたことを確認している。</p> <p>第4章では、少数の広告候補の選択問題を機械学習におけるマルチラベル分類問題で候補ラベル数が極めて多い場合として捉え、これに対する高精度・高速な方法を提案している。異なるラベルに対して別々の分類器を学習する方法では、候補ラベル数に比例して計算量が増加するという問題があった。本研究では、低次元埋め込みに基づくk-近傍ベースであるSLEEC法をもとに、さらに特徴空間の分割とグラフを用いた近傍探索によって効率的な学習と予測を実現する2種類の方法を提案している。大規模なデータセットを用いた実験では、予測精度と速度の両面で改善したことを確認している。</p> <p>第5章では、ユーザの行動履歴から、推薦に有用なユーザの興味の表現を抽出する問題に取り組んでいる。特に、ユーザのウェブ閲覧の履歴からユーザの興味を低次</p>			

元で表現するために、自然言語処理で利用されている段落埋め込みの手法を、ウェブページを単語に、段落を閲覧履歴に対応づけることでこの問題に適用している。自然言語で記述された文章中では一般に単語の出現頻度と文章中での出現位置にはほぼ関連性がないのに対して、ユーザの閲覧履歴では低頻度のウェブページは履歴中の後半に出現しやすいという実データの分析から得られた知見に着想を得て、時間を遡る方向に埋め込み処理を行うことを提案している。オンライン広告に関連する実データを用いた実験によって、ユーザの興味をより精度よく予測できることを確認している。

第6章は結論であり、本論文で得られた成果を要約している。即ち本論文は、大規模なウェブサービス上で、スケーラブルな推薦システムを構築する際にあらわれる2つの課題に取り組んだものであり、膨大な候補の中から各ユーザに適した少数の対象を限られた時間内で選択するために、効率的な検索システムに適した形でモデルを設計するか、あるいはモデル内に検索インデックスを含めるアプローチが有効であること、また、ユーザの行動履歴から、推薦に有用なユーザの興味の表現を抽出するために、セッション内の行動の出現位置を考慮してモデルを設計することが有効であることを示したものである。推薦システムが一層社会に浸透するためには、本研究の成果を一層発展させ、機械学習と検索インデックスとのより深い融合の実現や、推薦結果の多様性・公平性を確保することが重要であるといった展望とともに本論文を結んでいる。

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し  
審査結果の要旨は日本語500～2,000字程度で作成すること。

(続紙 2)

(論文審査の結果の要旨)

本論文は、大規模なウェブサービス上で、スケーラブルな推薦システムを構築する際にあらわれる2つの課題、即ち、膨大な候補の中から各ユーザに適した少数の対象を限られた時間内で選択する問題とユーザの行動履歴から推薦に有用なユーザの興味の表現を抽出する問題に取り組んだものであり、得られた主な成果は次の通りである。

1. 極めて多数の広告候補の中から数百程度の適切な候補を効率的に取り出すための方法として、転置インデックスに基づく方法があるが、通常、広告やウェブページと、これらと対になる検索クエリは、含まれる単語を用いると高次元で疎な表現となることが多く、語彙の異なりから類似度を計算することが困難であった。そこで、本研究では広告などに含まれる単語と、クエリに含まれる単語を対応づける変換表をデータから学習する問題を最適化問題として定式化し、実際の広告配信システムに提案手法を適用することにより、既存手法と比較して広告クリック率や収益を上げることができることを確認した。また、学習された変換表からは意味的にも妥当な対応が得られたことを確認した。

2. 前述の、少数の広告候補の選択問題を機械学習におけるマルチラベル分類問題で候補ラベル数が極めて多い場合として捉え、これに対する高精度・高速な方法を提案した。異なるラベルに対して別々の分類器を学習する方法では、候補ラベル数に比例して計算量が増加するという問題があったが、本研究では、対象の低次元埋め込みならびに、特徴空間の分割とグラフを用いた近傍探索によって効率的な学習と予測を実現する2種の方法を提案した。大規模なデータセットを用いた実験では、予測精度と速度の両面で改善したことを確認した。

3. ユーザの行動履歴から、推薦に有用なユーザの興味の表現を抽出する問題、特に、ユーザのウェブ閲覧の履歴からユーザの興味を低次元で表現するために、自然言語処理で利用されている段落埋め込みの手法を適用した。自然言語で記述された文章と異なり、ユーザのウェブ閲覧履歴では低頻度のウェブページは履歴の後半に出現しやすいという実データの分析から得られた知見に着想を得て、時間を遡る方向に埋め込み処理を行うことを提案し、オンライン広告に関連する実データを用いた実験によって、ユーザの興味をより精度よく予測できることを示した。

以上、本論文は大規模なウェブサービス上で、スケーラブルな推薦システムを構築する際にあらわれる課題に、おもに機械学習の立場から取り組んだ結果をまとめたものであり、学術上・実応用上寄与するところが少なくない。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30年8月20日に実施した論文内容とそれに関連した口頭試問の結果、合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。更に、試問の結果の要旨(例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」)を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。  
要旨公開可能日: 年 月 日以降