

## 審査の結果の要旨

氏名 福馬智生

インターネットの普及に伴う情報過多の時代、オンラインプラットフォームにおけるコンテンツの質の自動推定技術に注目が集まっている。代表的な試みとして集合知に基づく試み、教師ありランク学習、推薦システムなどが挙げられる。一方で、それら技術には様々な形でバイアスが介在し、ユーザーの意思決定に影響を与えることが知られている。そこで本研究ではそれら既存の試みにおけるバイアスの影響を取り除くことを目的とする。

本研究は大きく三部構成になっており、第一部は個人の投票行動における認知バイアスによる影響の除去、第二部は信頼性の観点から個人の意見をいかに集約するか、第三部は推薦システムにおけるアルゴリズムバイアスの緩和がテーマになっている。以下それぞれの概要を述べる。

第一部では、複数の認知バイアス情報(e.g., コンテンツの表示位置, 投稿者の評判, 他者による評価)が存在するサイトにおいて、バイアスのかかったユーザーのクリックデータから認知バイアスを除去した形でのコンテンツの有用性をどのように推定するかについて学習フレームワークの提案を行った。本研究ではユーザーの投票行動とコンテンツの有用度の関連性についてモデリングを行い、最後に、認知的要因の相互関係を定量化し、モデルの挙動について解釈を行った。

第二部では、既存レビュー信頼性アルゴリズム手法 REV2 によって算出されるレビューの信頼性に基づく集合知を実データを用いて検証した。結果として「やらせレビュー」や「アンチレビュー」といった信頼性の低いレビューの影響を軽減し、平均といった従来の集約方法よりも有識者に近い結果が得られた。加えてアルゴリズムの算出するスコアに対する解釈性の向上を行う。その結果、過小・過大評価されているモノについて特徴を解釈可能な形で抽出した。

第三部では、既存推薦システムにおける二つの大きな課題である疎なデータからの学習、不確実性の推論を可能にする学習フレームワーク **MetaCF** の提案を行った。提案フレームワークはメタ学習の一種である **Neural Processes** に基づいており、実質的に既存のフレームワークに手を加えることなく適用可能な点で優れている。本提案手法により履歴数が少ないユーザー・アイテムへの高精度な嗜好のモデリングが可能になり、また不確実性に基づく推薦によりフィルターバブルの解消への応用可能性が考えられる。

以上の通り、本研究によって、オンラインプラットフォームにおける根本的な疑問である、コンテンツの有用性をどのようにして測定するかを明らかにされた。

論文には、新規性、有用性が示されており、博士論文としての基準と満たされていると判断され、本論文は博士（工学）の学位請求論文として合格と認められる。