

CGM を利用した情報推薦

中田 瑛人

情報過多と呼ばれる現代、Web 上で自分の求める情報を探し出すことには困難が伴う。そのような状況で情報検索とともに研究されてきた技術のひとつに情報推薦がある。情報推薦はユーザの要求や嗜好を機械的に推定し、それらを満たす情報を推薦する技術であり、主に協調型、内容ベース型、知識ベース型の 3 つのモデルに分けられる。しかし、協調型にはアイテムの内容を考慮できないという課題が、内容ベース型には自明な推薦が行われやすいという課題が、知識ベース型にはユーザの負担が大きいという課題が存在する。

本研究では、それらの課題に対して協調型推薦を基本としながらも、アイテムの内容を推薦に反映する方式の開発を試みる。本研究で注目したのは、Consumer Generated Media(CGM)であるブックログと、そこで見られるタグである。CGM とは、インターネットで消費者によって内容が生成されるメディアであり、ブックログは図書をメインとするアイテムをユーザがレビューやタグを付与して評価できるレビューサイトである。ユーザの嗜好、アイテムの内容に対する評価を含むこの「タグ」を協調型推薦に利用した新しい推薦モデルをタグベース協調型推薦と呼ぶ。

タグベース協調型推薦では、ユーザがブックログに登録したアイテムから嗜好タグを抽出し、比較することで類似ユーザを見つける。嗜好タグ抽出の基準として、2 種類の尺度を導入しシステムを実現した。タグの全アイテムでの出現頻度の逆数に対数をとった値(IDF)を用いたシステム(1)と、ユーザの登録アイテム内でのタグの出現頻度(TF)を用いたシステム(2)の 2 種類である。

新しく提案した推薦モデルの有効性を評価するため、(1)と(2)のタグベース協調型推薦システムに加え、従来の協調型推薦を実現したシステム(3)を実現し、ブックログのユーザデータを使った推薦シミュレーションを行った。実験では、ユーザの登録アイテムをクエリ用とテスト用に分割し、クエリ用のアイテムを入力とした推薦結果アイテムとテスト用アイテムを照合した際の再現率と適合率をシステムの評価の値とした。

実験の結果、再現率では(2) > (3) > (1)という結果に、適合率では(3) > (1) > (2)という結果になった。嗜好タグの抽出に TF を用いた(2)では、嗜好タグに高頻度のタグがよく現れ、類似ユーザが多く計算されたことで、再現率は上がった。しかし、その類似ユーザに嗜好が近いとは言えないユーザが多く含まれてしまったことで適合率が下がってしまった。タグの抽出に IDF を用いた(1)では、ユーザにとって重要度の高いと思われるタグが嗜好タグに現れ、より嗜好の近い類似ユーザを計算することができたため、適合率が上がったと思われるが、類似ユーザの候補を一人も獲得できなかったユーザが多く、再現率は低かった。

(指導教員 阪口哲男)