

5U-5 組合せフィルタリング方式を用いた情報推薦システム

有吉 勇介

NEC ヒューマンメディア研究所

1.はじめに

近年、インターネットが急速に普及し、まさに情報洪水といわれる事態を招いている。情報を受け取る側にとっては、大量の情報のなかから自分の関心に合ったものだけを選別することが必要になっており、情報を発信する側にとっては、相手の関心にマッチする形で情報を提供することが、サービス価値を高めるための重要になってきている。そのため利用者の関心を学習し、関心に合った情報だけを推薦する情報推薦システムが注目されている。

しかし従来の情報推薦システムでは、利用者の関心をシステムに伝えるために面倒なフィードバック操作が必要であった [1]。面倒な操作を無くすために、利用者の行動から利用者の関心を推定する研究も行われているが、行動から間接的に関心を推定すると推薦の精度が低下するという弊害が起ってしまう [2]。

利用者の「関心に合った情報を①面倒な操作を必要とせずに②高い精度で推薦して欲しい。」という情報推薦システムへの要求は、技術的にトレードオフの関係にある。本稿では、この両方の要求を満たすために筆者らが開発した情報推薦システムについて紹介する。

2.情報推薦システム

2.1.評価の推定

従来の情報推薦システムでは、利用者の関心をシステムに伝えるために、推薦された情報が利用者にとって有用だったか否かを5段階評価で返すなどの面倒なフィードバック操作が必要であった。この面倒な操作を無くすために、情報を読むのに費やした時間（注視時間）から利用者の情報に対する関心度を推定する研究も存在するが、関心を行動から間接的に推定すると推薦の精度が低下するという弊害が起ってしまう。

An Information Recommender System Using Hybrid Filtering Method
Yusuke Ariyoshi
Human Media Research Laboratories, NEC Corporation
E-mail: ariyoshi@hml.cl.nec.co.jp

この弊害は、注視時間が利用者の情報に対する関心度を正確に反映していないことに原因があり、利用者の関心をより直接反映した利用者行動を利用すれば、弊害を低減できると考えられる。そこで本稿の情報推薦システムでは、 unnecessary 情報を削除したり、特に関心のある情報を保存するなどの利用者が推薦された情報を整理する操作から情報に対する関心度の推定を行う。

2.2.情報フィルタリング方式

本稿の情報推薦システムでは、高い精度の推薦を提供するために、論文[3][4]の方式を組合せた情報フィルタリング方式を採用している。情報フィルタリングは、利用者の関心の学習と評価の予測を行う技術であり、情報推薦の基本技術である。

従来の情報フィルタリング方式は、大きくソーシャル情報フィルタリング方式(SIF)と内容ベースフィルタリング方式(CBF)の2種類に分けることができる。SIFはユーザAと関心が類似した別のユーザBが高い評価をしている情報を、ユーザAにも薦めるというものである。CBFでは、各ユーザの関心を表現するキーワードベクトルと、情報内容を表すキーワードベクトルとをマッチングする。

本稿の情報推薦システムは以下のようにSIFとCBFを効果的に組合せることで高精度の推薦を提供する(図1)。

評価変換：ここでは、評価値を段階尺度から順位尺度へ変換を行う。従来の方式では利用者が情報に与えた評価

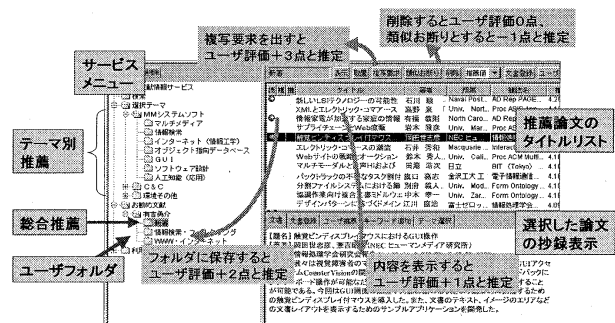


図1：画面例

値をそのまま使用していた。しかし、評価値が同じであっても評価が厳しい利用者と緩い利用者ではその意味合いは異なるはずである。そこで、評価値を順位尺度へ変換を行うことによって評価値は各利用者の評価の付け方に応じた数量に変換される。

推薦統合：ここでは CBF と SIF の推薦を統合する。このとき、CBF と SIF それぞれからの推薦の信頼性を考慮することで、推薦の精度を高めている。推薦の信頼度は、推薦に用いたデータの量（例えば CBF では学習したキーワード数、SIF では類似利用者の人数など）が多いほど高くなる。そこで推薦に用いたデータ量からそれぞれの推薦の信頼度を推定し、推定した信頼度を重みとして両者の推薦の統合を行っている。

学習強化：ここでは、CBF の単語重要度の学習を SIF 結果で強化している。つまり、単語重要度を、その利用者の評価済み情報に加えて、SIF による予測評価値を評価値の代わりにすることでその利用者が未評価の情報からも学習している。

3. システム構成

この節では本稿の情報推薦システムの応用として実装した論文クリッピングサービスについて説明する(図2)。

①この論文クリッピングサービスでは、各利用者の関心に合わせたお薦め論文のタイトル一覧が、各利用者ごとのホームページに提示される。利用者は、タイトルだけでなく論文概要を開いて見たり、必要な論文はフォルダに分類整理したり、不要なものは削除するなどの操作が行える。情報推薦システムは、このような利用者の操作から各々の論文に対する利用者の評価を推測し(図3)。

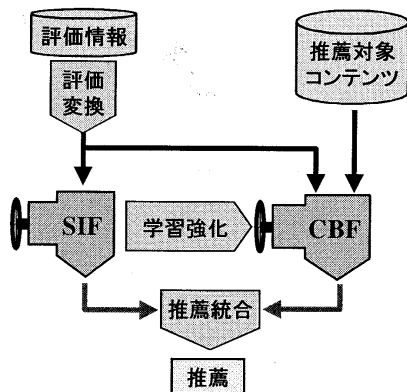


図2: 情報フィルタリング方式

その評価に基づいて利用者プロフィールを更新する。このようにして次第に利用者の関心をより正確にとらえた論文提示する。

②論文情報は科学技術振興事業団の JICST データベースを利用し、週一回システムに登録される。新着論文情報は、キーワード抽出を行い、文書インデックスが追加される。

③新規利用者の登録があると、その利用者が執筆した論文の情報を社内文書データベースから取得し、キーワード抽出を行って、初期の利用者プロフィールを生成する。

4. おわりに

本稿で紹介した情報推薦システムは、利用者の推薦情報を整理する操作から利用者の情報に対する評価の推定を行う。これにより、面倒なフィードバック操作が不要であり、従来の注視時間より高い精度で評価が推定できる。また、SIF と CBF を効果的に組合せた情報フィルタリング方式を用いることで、精度の高い推薦を提供する。また、本情報推薦システムの応用として開発し、社内内で運用を実施している論文クリッピングサービスの紹介を行った。

参考文献

- [1] J.A.Konstan, et al., "GroupLens: applying collaborative filtering to Usenet News", CACM, Vol.40, No.3, 1997.
- [2] 鎌原ら, "利用者情報を用いた電子ニュース記事評価機構の試作", 第22回 jusUNIX シンポジウム予稿集, 1993
- [3] 有吉, "組合せ情報フィルタ方式の信頼度を用いた精度改良", 情処研報, FI-53-4, 1999年
- [4] 有吉, "利用者毎に評価尺度変換を行うソーシャル情報フィルタ方式", 情処58 全大, 5R-8, 1999年
- [5] "Special Section: Recommender Systems", CACM, Vol.40, No.3, pp.56-89, Mar 1997.

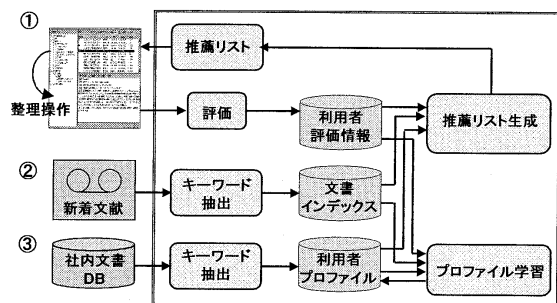


図3: システム構成