

ユーザ嗜好に基づく ニュース記事の推薦に関する研究

檜山武浩[†] 田中成典^{††} 杉之原亮[†]

関西大学大学院総合情報学研究科[†] 関西大学総合情報学部^{††}

1. はじめに

近年、インターネットの普及に伴い、誰もが Web から様々な情報を取得可能になった。特に、Web のニュース記事は、新聞と比較して即時性に優れている利点から多くの人によって収集されている。さらに、更新情報を XML ベースのフォーマット文書で配信する RSS (RDF Site Summary) の登場により、ユーザはニュース記事の効率的な収集が可能になった。しかし、RSS を配信するサイトの増加により、収集した RSS の情報が氾濫し、情報収集効率が低下する問題が生じている。この問題に対して、ユーザ嗜好に基づいた情報推薦技術[1][2]が注目されている。この技術は、パーソナライゼーション[3]とも呼ばれ、ユーザの行動履歴などからユーザ嗜好を表現したユーザプロフィールを構築し、それに基づいて情報を推薦する技術である。情報推薦技術においては、ユーザプロフィールの精度が推薦の精度に大きく影響する。ユーザプロフィールにおける精度向上の研究では、遺伝的アルゴリズムを用いることにより、ユーザプロフィールを構築する手法[4]を提案している。しかし、この手法により作成したユーザプロフィールでは、閲覧履歴などのユーザが明示した情報の範囲でしか推薦できない。そのため、推薦できるニュース記事リストの内容が閲覧履歴から抽出した内容に偏ってしまい推薦する記事内容の多様性が低下してしまう。そこで、本研究では、ユーザの行動履歴に加えてユーザ嗜好と類似したネットユーザ（ブロガーやソーシャルネットワーキングサービスのユーザなど）の行動履歴を考慮することにより、ユーザ嗜好を幅広く網羅した多様性のあるニュース記事リストを推薦する手法を新たに提案する。

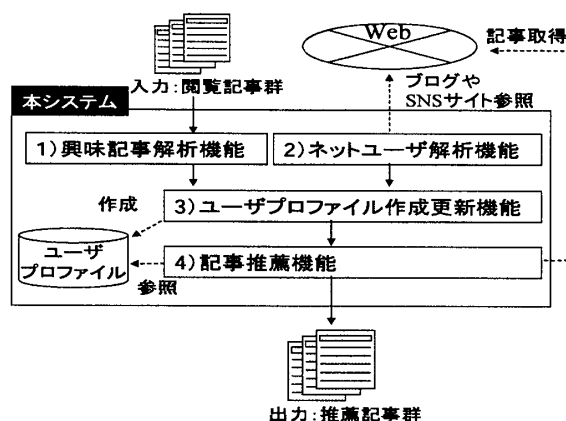


図 1 本提案手法の概念

2. 本研究の概要

本研究では、ユーザ嗜好に基づいた多様性のあるニュース記事リストを推薦することを目的とする。本提案手法の概念を図 1 に示す。まず、「興味記事解析機能」と「ネットユーザ解析機能」により、本システムを利用するユーザとネットユーザの興味記事を解析する。前者は、ユーザの記事閲覧時間を用いて興味記事を解析し、後者は、ブログ記事やソーシャルネットワーキングサービスの日記内でのニュース記事に対するリンクの有無から興味記事を解析する。次に、「ユーザプロフィール作成更新機能」により、「興味記事解析機能」と「ネットユーザ解析機能」で解析した興味記事からユーザプロフィールを作成し、過去のユーザプロフィールを更新する。ユーザプロフィールは、実数値を個体として扱える実数値遺伝的アルゴリズム[5]を用いて、興味語の最適な興味度を探索することで作成する。個体は、特徴語に対する興味度を実数ベクトルで表現するユーザプロフィールとする。最後に、「記事推薦機能」では、「ユーザプロフィール作成更新機能」で構築したユーザプロフィールを参照して、ユーザの興味度が高い順番に Web から取得したニュース記事を並び替えて提供する。興味度は、ユーザプロフィールとニュース記事との類似度として算出する。

Research for Recommending News Articles Based on User's Preference

[†]Takehiro Kashiyama, Ryo Suginoara

Graduate School of Informatics, Kansai University, 2-1-1 Ryouzenji-cho, Takatsuki-shi, Osaka, 569-1095, Japan

^{††}Shigenori Tanaka

Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-cho, Takatsuki-shi, Osaka, 569-1095, Japan

3. 実証実験

本研究で提案した手法の有効性を示すため、ニュース記事の推薦精度について既存手法と比較実験して検証する。既存手法は、ネットユーザの嗜好を考慮せずに遺伝的アルゴリズムを用いて構築したユーザプロファイルに基づく推薦手法とする。遺伝的アルゴリズムでは、世代数を 100000 世代、親個体数を 500 個体、交叉回数を 50 回とした。

(1) 実験内容

実験では、両手法ともニュース記事の閲覧作業を 1 週間行い、システムにより提供されたニュース記事リストからユーザにとって興味があるニュース記事の割合（正確性）と推薦されたニュース記事リストにおける内容の類似度（多様性）を求めた。「正確性」は、情報検索システムの抽出精度を表す指標として一般的に利用されている適合率、再現率と F 値により評価する。適合率、再現率と F 値は、各手法により提供された上位 20 件のニュース記事とユーザの興味記事集合から算出した。興味記事集合は、日ごとに取得した全ニュース記事からユーザが興味を持った記事を目視で確認して集計した。

「多様性」は、システムが推薦したニュース記事リストにおける内容の多様性を測定する指標であり、各手法により提供された上位 20 件のニュース記事間における記事内容の類似度の合計により評価した。記事間における内容の類似度は、高いほど記事内容が偏っており、多様性が低いことを示す。本実験では、20 件の記事内容の類似度の合計が「多様性」の評価値となるため、全て同じ記事内容の場合、「多様性」の評価値は、190 となる。

表 1 実験結果

	正確性			多様性
	適合率	再現率	F 値	
本提案手法	0.75	0.67	0.71	5.43
既存手法	0.80	0.71	0.75	6.27

(2) 結果と考察

実験結果を表 1 に示す。本実験では、1 週間で取得した全ニュース記事数は 1064 件、その中でユーザの興味がある全ニュース記事は 157 件であった。表 1 の結果から「正確性」において本提案手法では、140 件の推薦記事のうち 105 件がユーザの興味記事であった。一方、既存手法では、140 件の推薦記事のうち 112 件がユーザの興味記事であった。「多様性」において本提案手

法では、5.43 となり、既存手法では、6.27 となった。このことから、本提案手法は、「正確性」では適合率と再現率とも既存手法より若干低い結果となったが、「多様性」では、既存手法より推薦記事リストの内容に偏りがなく幅広い内容の記事を推薦することができた。本提案手法の正確性が既存手法より若干低くなった要因としては、遺伝的アルゴリズムを用いてユーザプロファイルを作成するときに本提案手法では、ユーザの興味記事に加えて類似ネットユーザの興味記事を考慮して個体を生成するため、遺伝子数が既存手法よりも多くなり最適なユーザプロファイルに収束しなかったことが考えられる。

4. おわりに

本研究では、ユーザ嗜好と類似したネットユーザの行動履歴を考慮してユーザプロファイルを構築することでユーザ嗜好を幅広く網羅した多様性のあるニュース記事リストを推薦する手法を提案した。実験の結果、本研究の有用性を証明できた。ただし、本研究では、ニュース記事を一覧表示で提供しているため、提供されたニュース記事の内容を直感的に把握できない。今後の課題は、ユーザ嗜好とニュース記事の関係を可視化し、全てのジャンルのニュース記事を俯瞰することで、情報収集効率のさらなる向上を実現させることである。

参考文献

- [1]Tomonari Kamba, Hidekazu Sakagami, Yoshiyuki Koseki : ANATAGONOMY: a personalized newspaper on the World Wide Web, International Journal of Human-Computer Studies, Vol.46, No.6, pp.789-803, 1997.6.
- [2]土方嘉徳：情報推薦・情報フィルタリングのためのユーザプロファイリング技術, 人工知能学会誌, 人工知能学会, Vol.19, No.3, pp.365-372, 2004.5.
- [3]Riecken Doug : Personalized Views of Personalization, Communications of the ACM, Association for Computing Machinery, Vol.43, No.8, pp.26-158, 2000.8.
- [4]柳本豪一, 藤中透, 吉岡理文, 大松 繁：情報フィルタリングにおける遺伝的アルゴリズムを用いたユーザプロファイルの作成, 電気学会論文誌 C, 電気学会, Vol.121-C, No.7, pp.1277-1282, 2001.7.
- [5]小野功, 佐藤浩, 小林重信：単峰性正規分布交叉 UNDX を用いた実数値 GA による関数最適化, 人工知能学会誌, 人工知能学会, Vol.14, No.6, pp.1146-1154, 1999.11.