

「情報地図」による情報獲得の支援に関する研究

川瀬宏一郎† 篠田陽一† 落水浩一郎†

†北陸先端科学技術大学院大学 情報科学研究科
〒923-1292 石川県能美郡辰口町旭台 1-1

E-mail: {k-kawase, shinoda, ochimizu}@jaist.ac.jp

あらまし

本稿では、協調的情報フィルタリングに基づく情報獲得支援システムについて述べる。本システムの設計方針は、個人が保有する情報体系を「情報地図」と呼ぶグラフ形式にて視覚化し、各ユーザの情報地図をシステムが合成することにより、他ユーザの知識を参照可能にするものである。また、プロトタイプシステムを実装し、評価実験を行った。その結果と考察について報告する。

Information Acquisition Support by Information Map

Koichiro KAWASE† Yoichi SHINODA† Koichiro OCHIMIZU†

†School of Information Science,
Japan Advanced Institute of Science and Technology,
1-1, Asahidai, Nomi-gun, Ishikawa, 923-1292 Japan

E-mail: {k-kawase, shinoda, ochimizu}@jaist.ac.jp

Abstract

In this paper, we describe principles, system, and evaluation of information acquisition support system based on collaborative information filtering. The purpose of our system enable users to use others' knowledge by merging the information each has. The knowledge representation named Information Map is a graph, each node represents the information unit and an edge represents the navigation link to get useful unknown information. We evaluated effectiveness of our system through experiment using the prototype system.

1 はじめに

ソフトウェア設計グループなどの、目的を共にした少人数で構成されている集団にとって、作業の遂行に必要な知識を効率良く獲得するための技術は、重要な課題となっており、ネットワーク時代の知的生産活動においては、さらに発展させるべき事柄である。このための一つの有力な手段として協調的情報フィルタリング (Collaborative Filtering) がある。この方法の基本的立場は以下の通りである。

「何らかの興味を共有している人同士で、互いの知識を参考にして情報をふるいにかける」

情報獲得活動を個人の知識による限定された形で繰り返すことは、その人の知識の広がりや硬直化を招くことになりかねない。協調的フィルタリングは、そのような問題を回避する一つの手段として期待できる。

2 協調的情報フィルタリングシステム

2.1 既存システムの特徴

協調的フィルタリングを利用したシステムには、大きく分けて下記の2つの方式がある [1]。それぞれの方式において、既に幾つかのシステムが存在しており、その特徴も合わせて記す。

1. 相関に基づくフィルタリングシステム
ユーザが行った評価の履歴により、嗜好が類似した集団を割り出し、その集団内での評価を参照し合う方式である。この方式のシステムでは、ユーザの増加により、フィルタの機能を向上させることができる。例えば、Netnews の記事に対し、大人数のユーザの存在を前提としてフィルタリングを実現するシステムに、GroupLens [2] がある。
2. 推薦に基づくフィルタリングシステム
自分が得た情報に注釈を付けて、他ユーザに推薦することができる方式である。この方式のシステムの具体的な例としては、Pointer [3] があるが、コメント付けの作業は煩雑なものになりがちである。



図 1: 「既知のこと」と「未知のこと」のつながり

2.2 協調的フィルタリングシステムに関する要請

第 2.1 節における既存システムの評価を踏まえて、協調的フィルタリングの望ましい特質を以下のように設定した。

- 必ずしも大人数での利用を前提条件としない設計方針を採用したい。
- 他者に情報を利用させるための手段を考慮する必要がある。例えば、情報に対して、有用度の点数付け程度ならまだしも、自然言語による注釈付けをするとすると煩わしい。
- 様々な情報ソースから情報を獲得してきて必要があり、特定の情報ソースに対してしかフィルタリングの機能を提供しない設計は望ましくない。
- 情報の表現形式を工夫する必要がある。有用だと思われる情報を、単にリストによって提示するだけでは、各情報がユーザにとってどのような意義を持つのかが分かりにくい。カテゴリ分類による情報の提示も考えられるが、欲しい情報がユーザの予想していたカテゴリ内に収まっているとは限らない。

3 システムの設計方針

本章では、第 2.2 節で述べた要請を考慮しつつ、新たなフィルタリングシステムの設計方針を述べる。

他者から提供された新しい情報と自身が既に保有している情報との関連が、わかりやすい形で表されるのが望ましい (図 1)。なぜならば、その関連を示すことで、提供された新しい情報が持つ意義を示唆することが可能であると考えられるからである。

この方針を具体化するために、本研究では「情報地図」と呼ぶ形態で情報を表現する (図 2)。情報地図は、個人が持っている情報間の関係をグラフによっ

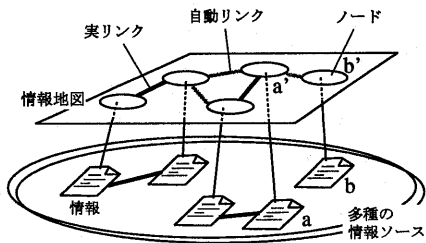


図 2: 情報地図

て明示的に視覚化したものである。情報地図の特徴を以下に挙げる。

情報地図は、個人が持っている情報間の関係をグラフによって明示的に視覚化したものである。情報地図の特徴を以下に挙げる。

1. 情報地図のノード

- 「WWW のコンテンツ (図 2 中の a)」、
「個人的なファイル (図 2 中の b)」といった、異なる情報ソースに存在している情報を一元的に表示する (図 2 中の a', b')。

2. 情報地図のリンク

- WWW 上の URL によるリンクのように、現実世界で明示化されている情報の関係を反映する (図 2 中の「実リンク」)。
- 実際には関係付けられていない情報に対しても、その情報を解析することにより、自動的にリンクを張る (図 2 中の「自動リンク」)。tf*idf 法¹を用いてキーワードの重み付けを行い、重要キーワードの重なり具合によって自動リンクを生成するか否かを判定する。
- ユーザの手によるリンクの編集も可能である。実リンク、自動リンク共に存在しない情報間にも、ユーザの主観によりリンクを張ることができる。また、生成された自動リンクを消去することもできる。

本システムは、各ユーザ毎に情報地図を生成し、合成する。ユーザは、合成された情報地図を参照することで、新たな情報を知ることができる。

¹ 「文書 d_j における単語 t_i の出現頻度」と、「全文書における単語 t_i の特殊度」から、「文書 d_j における単語 t_i の重要度」を求める手法 [4]

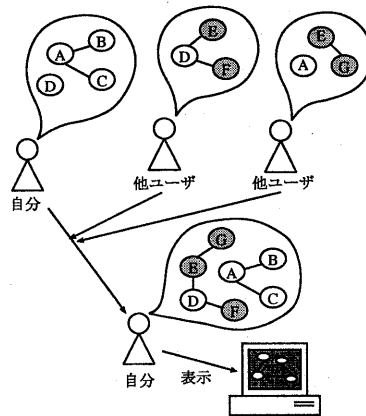


図 3: 情報地図の合成

4 プロトタイプシステムの概説

第 3 章で述べたシステムの設計方針に基づき、プロトタイプシステムを実装した。プロトタイプシステムの利用の流れと構成の概略を、図 4 を基に述べる。

プロトタイプシステムの利用の流れは以下の通りである。

1. 本システムの各ユーザは、日頃から個人的に情報獲得活動を行っている。そして、その活動によって日々更新されるシステム用の設定ファイルを各自持っており、そこには、
 - 自分で見つけてきた情報の URL
 - 他ユーザが持っている情報の URL
 などが記録されている。
2. ユーザは、自分の端末上でシステム用のコマンドを入力すると、システムは自分の設定ファイルを読み込み、その内容に応じた情報地図用のデータを生成してくれる。情報地図データが更新されたかどうかは、端末の画面上に表示される。
3. ユーザは、情報地図データが更新されたのを確認したら、WWW ブラウザの Netscape を起動し、情報地図システムを提供している URL を open する。すると、java アプレットがロードされる。そのアプレットには、情報地図が表示されている (図 5)。

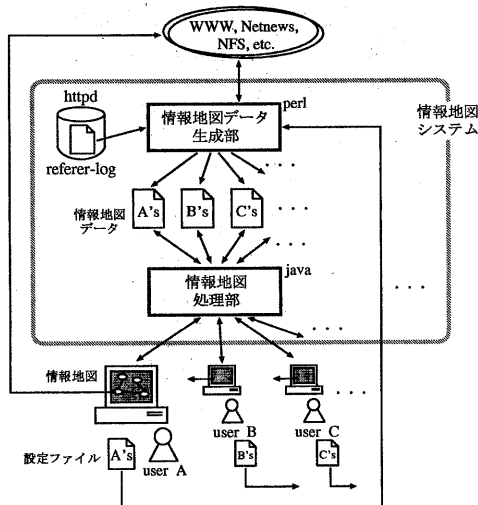


図 4: プロトタイプシステムの構成

4. 彼にとって未知の情報として提示されているノードをダブルクリックすることによって、その情報にアクセスすることができる。
また、自動リンクの修正を行うことができ、その操作によっても、情報地図データが更新される。

図 4中の「情報地図データ生成部」は、情報地図用のデータを生成するためのバックエンドであり、プログラミング言語 Perl²で記述してある。上記の利用シナリオの 2. に相当する部分を処理する。「情報地図処理部」は、情報地図データを処理し、グラフィカルな情報地図表示を行い、さらに、ユーザからの反応を処理するためのフロントエンドである。こちらは Java 言語³で記述してある。上記の利用シナリオの 3.、4. の部分を受け持っている。本システムは、一台のコンピュータ上に実装されている。さらに、そのコンピュータは、WWW proxy サーバも兼ねている。本システムのユーザは、普段の WWW を利用する際にも、このコンピュータを proxy サーバとして利用することで、ユーザの WWW のアクセス履歴がシステム内に保存される。

²version 5.002

³Java Development Kit ver.1.1.3 を使用

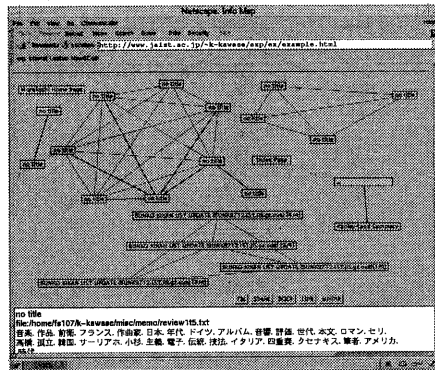


図 5: 情報地図アプレットの例

5 評価実験

5.1 評価の目標

情報地図による協調的フィルタリングの有効性を確認する実験をプロトタイプシステムを利用して行った。評価の目標は以下の 2 点を確認することにあつた。

評価目標 1 他ユーザの情報地図を合成することで有用な情報が得られるか?

評価目標 2 自動リンクを用いたグラフ表示は、有用な情報を得るのに有効か?

5.2 実験内容

被験者は、本大学院の院生 3 名である。彼らに、「自分の計算機環境のカスタマイズのために、有用な情報をインターネットから得る」という情報獲得に関する目標の下に、実際にプロトタイプを利用してもらった。

実験の流れを以下に示す。

1. 被験者：各自が普段行っている方法によって情報を収集する。これは随時行われる。
実験者：被験者に、情報地図を利用してもらう時間帯を案内しておく。
2. 被験者：各自が自分で手に入れた情報を実験者に知らせる。その後、情報地図を呼出し、参照する。
実験者：被験者からの情報を受取り、被験者が情報地図を呼び出す前に、情報地図データを更新しておく。

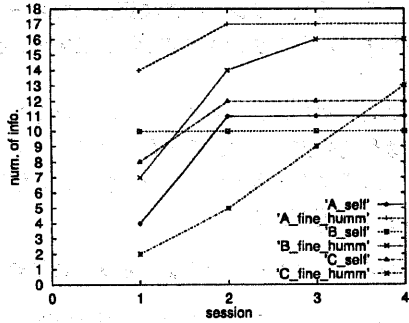


図 6: 情報の累積結果

3. 被験者: 各自が参照している情報地図から得られる情報を評価する。

実験者: 各被験者が行った評価データを収集し整理する。

4. 被験者: 3. の作業が終わったら、1. の作業に戻る。
 実験者: 被験者に 1. の作業に戻るように知らせる。

2., 3. は 1 日 1 回連続して行われる。これを 1 セッションとしてカウントする。また、1. での情報の増加が停止することによって 2. で参照される情報地図が更新されなくなる、という状況が 2 回連続したら、実験を終了することにした。結果的には、3 日目には 1. での情報の増加はストップし、4 セッションで実験は終了した。

5.3 情報地図の合成による効果

第 5.1 節で述べた評価目標 1 「他ユーザの情報地図を合成することで有用な情報が得られるか?」について、2 つの観測を基に考察する。

5.3.1 新しくかつ不要でない情報の累積結果

第 5.2 節で述べた 1. での作業による情報の累積数と、各自が「自分にとって新しくかつ不要でない」として評価した情報の累積数を図 7 に示す。図中の 'A_self' は、被験者 A が自分で見つけてきた情報の累積を表し、'A_fine_humm' は、彼にとって新しくかつ不要でない情報の累積を表している。

この結果から、システムの利用によって、各被験者は、「新しくかつ不要ではない情報」を増やしていることが定性的に言える。

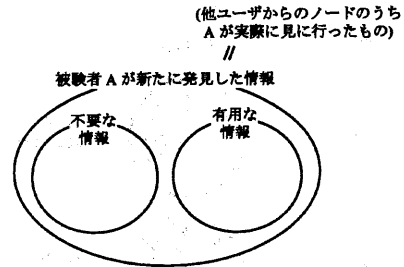


図 7: 有用な情報と不要な情報のモデル

5.3.2 情報の有用度と不要度

被験者が新たに発見した情報のうち、自分にとって有用だった情報の割合と、不要だった情報の割合を求める。これらの数値は、以下の式で求める。なお、「被験者 A が新たに発見した情報の数」は、「他の被験者から提供されたノードのうち、A が実際に見に行った数」ということになる。

有用度 =

$$\frac{\text{被験者 A にとって、新しくかつ有用な情報数}}{\text{被験者 A が新たに発見した情報数}}$$

不要度 =

$$\frac{\text{被験者 A にとって、不要な情報数}}{\text{被験者 A が新たに発見した情報数}}$$

有用度と不要度の算出結果を図 8 に示す。

この結果から、自分にとって新しい情報を見に行くとき、その情報は有用である可能性が高いことが言える。

以上の 2 つの観測結果から、本システムの利用により、自分にとって新しくかつ有用な情報を、確実にしかも効率良く獲得することができるという。

5.4 グラフ表示の有効性

第 5.1 節で述べた評価目標 2 「自動リンクを用いたグラフ表示は有用な情報を得るのに有効か?」について考察する。そのために、図 5.4 に示す数値 F と B の比較を考えた。

しかし、この数値の比較からは有意な傾向は見出せなかった。その理由を考察するために、被験者に対しアンケート調査を行ったところ、以下の指摘があった。

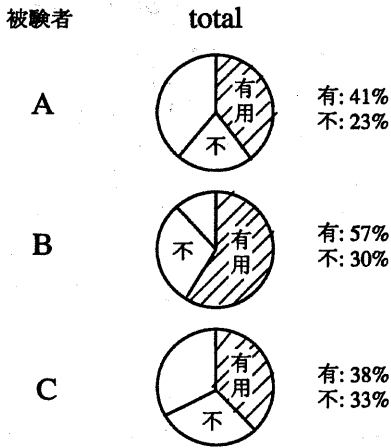
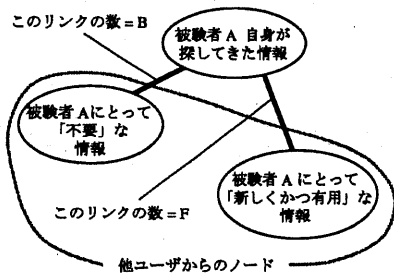


図 8: 有用度と不要度の結果



- リンクの数が多いのでリンクの意義を掴みづらい
- リンクの表示が煩雑になり使いづらい

6 まとめと今後の課題

現在の段階で、情報地図によるフィルタリング方法を提案し、システムの利用により、ユーザ単独で行う情報獲得よりも効率良く有用で新しい情報を手にすることができることを示した。しかし、情報間のリンクの有効性については、

- リンクの数が多すぎる場合は、リンク一つ一つの意義が薄れてしまう
- ユーザインタフェースの不備により使いづらい

という問題が残された。この問題点に対処するために、まずはリンクの自動生成の確度を向上させる必

要がある。また、リンクへの注釈としてキーワードを付加するなどのリンクの表示法を工夫する必要がある。

さらに、以下のようなドメインにおけるシステムの有効性を評価したい。そのためには、人数、利用目的、利用期間の各パラメータを変えた運用実験を実施する必要がある。

- あるメーリングリスト参加者間での情報交換
- Netnews の記事を対象にした情報収集
- ソフトウェアプロジェクト全体またはサブグループにおける情報共有

謝辞

本研究は、文部省重点領域研究「発展機構を備えたソフトウェア構成原理の研究」(課題番号 09245104)の援助の下に実施された。

参考文献

- [1] 岩山登. いかにしてネットワークから情報を得るか: キーワードによる探索と、コミュニティからの示唆. 「インターネットと情報検索」講習会資料 (日本ソフトウェア科学会), pp. 75-91, 1996.
- [2] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Gropulens: An open architecture for collaborative filtering of netnews. In *Proceeding of the ACM 1994 Conference on Computer Supported Cooperative Work*, pp. 175-186, 1994.
- [3] David Maltz and Kate Ehrlich. Pointing the way: Active collaborative filtering of netnews. In *Proceeding of the ACM Conference on Human Factors in Computing Systems, CHI'95*, pp. 202-209, 1995.
- [4] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, Vol. 18, No. 11, pp. 613-620, Nov. 1975.