

個人適応型情報検索システム
— 個人の興味を学習する階層記憶モデルと
その協調的フィルタリングへの適用 —

野美山 浩† 紺谷 精一‡ 渡辺 日出雄† 串間 和彦‡ 堤 泰治郎†

†日本アイ・ビー・エム株式会社 東京基礎研究所

‡日本電信電話 (株) NTT 情報通信研究所データベース研究部

近年のネットワークの発達により、多種多様で膨大な情報にアクセス可能となっている。このような情報氾濫の状況においては、大量の情報の中から個人の興味に合った情報を自動的に選択する機能が重要となる。本稿では、まず、個人の興味を学習するために、階層的な構造を持つ記憶モデルを定義し、読んだ情報の内容を解析し、それを記憶に反映させることで、個人に適応する情報検索を可能とする方式を提案する。さらに、その個人記憶モデルを、集団の中で他の人の評価に基づいて情報を推薦する協調型フィルタリングに適用する方法について述べる。

Personalized Information Navigator :
the Hierarchical Memory Model for Learning Users' Interests
and Its Application to Collaborative Filtering

Hiroshi Nomiyama†, Seiichi Kon'ya†, Hideo Watanabe†,

Kazuhiko Kushima†, Taijiro Tsutsumi†

†IBM Research, Tokyo Research Laboratory

1623-14 Shimotsuruma, Yamato, Kanagawa 242, Japan

Phone: 81-462-73-4563 E-Mail: {nomiyama,watanabe,tsutsumi}@trl.ibm.co.jp

‡Database Systems Laboratory

NTT Information and Communication Systems Laboratories

1-2356 Take Yokosuka-Shi, Kanagawa 238-03, Japan

Phone: 81-468-59-8611 E-Mail: {konya,kushima}@syrinx.ntt.jp

Huge amounts of varied information, such as electronic mail and network news, have become accessible as a result of the development of networks in recent years. This flood of information increased the importance of retrieving information from large amounts of available data and categorizing it according to a personal viewpoint. First, we define a hierarchical memory model that accumulates data on a user's interests through his or her retrieval activities and reflects them in the memory in order to allow information to be retrieved according to a personal viewpoint. And we applied this model to collaborative filtering, which is used to recommend useful information based on other users' rating.

1 はじめに

電子メールや電子ニュース等の爆発的拡大により、個人がアクセス可能な情報は飛躍的に増大しつつある。即ち、多種多様な情報提供者がネットワークを介して提供する情報の氾濫により、情報利用者にとっては自分に必要な情報を取捨選択するのがますます困難になっている。これらの情報洪水に効果的に対処する手段として情報フィルタリング [1, 2, 3, 4] の研究がなされている。これらは個人の興味をルール、もしくは事例して蓄え、新しい情報の出現に対し、蓄積した学習内容をもとに選択を行うものである。

我々はまず、個人の興味を管理する枠組みとして、階層記憶モデルを考案し、そのモデルを新聞記事データベースに適用し、個人の興味を記憶中に、重み付けを変化させることで個人の興味に適合可能な検索方法を提案する。

さらに、この個人記憶モデルを協調的フィルタリングに適用する方法を提案する。協調的フィルタリング (social filtering あるいは collaborative filtering) [1, 2, 3, 4] は、個々の情報の中身を解析することによって情報を選別する手法 (content-based filtering) とは異なり、情報の中身を解析せずに、個々の情報に対する評価を利用することによって個人の興味に一致する情報を選別する手法である。

この手法は、中身が何であるかをまったく知らなくても可能なため、多様な情報の集合に対する情報選別に対して1つの有効な手段となる。従来、その特徴付けが困難であり、ほとんど有効な情報選別手段がなかった動画・静止画などのマルチメディア情報に対しても選別の手段を与える。

しかし、この手法には、以下に示すような欠点がある。

- ある程度の量のユーザからの評価が集まらなければ、有効な情報が得られない
基本的に他の人間が何らかの形でアクセスしたものしか得られない。しかも、興味の近い人間についてその情報が得られなくてはならない。
- 広い分野に適応した場合発散する恐れがある
現在報告されているのは、音楽、net news、など比較的狭い分野である。また、HOMR [1] に対するユーザのコメントから音楽的に広い趣味を持っている人には、余り良い推薦が来ていないようにも推察される。また、個人の興味は多様であるため、単に個人の興味が1つであるという前提では、違った興味に一致する情報が選別される恐れがある¹。

¹例えば、次のような例。「AさんとBさんはともに、「情報検索」に興味があるが、Aさんは釣りにも興味があり、その情報に対して高い評価を与えているものとするならば、従来の方法

- 何故それが推薦されたのかが直接的にはユーザに明確でない

HOMR では、artist や album にコメントを加えることができ、趣味が同じユーザのコメントのみを表示するような工夫もされているが、十分であるとはいえない²。

我々は、協調的フィルタリングの長所を活かしつつ、その欠点を解消するために、個人記憶モデルを適用することを試みた。本稿では、内容の解析が実用レベルである程度可能なテキストを対象とし、従来の内容に基づくフィルタリングと協調的フィルタリングを混合した形の新しいフィルタリング手法を個人記憶モデルを用いて実現する手法を提案する。

続く章で、まず、記憶モデルに対する要求条件について述べ、個人における記憶モデルについて定義する。次に、その個人の記憶モデルを協調的フィルタリングに適用する方法について述べる。

2 記憶モデルに対する要求条件

学習により情報の選択を行う場合、下記の条件が必要である。

学習の暗黙性 ユーザに学習させていると意識させることなく、ユーザの興味を学習できること。

分野依存性 ユーザの持つ複数の分野の興味を管理し、反映できること。

効率 大量の情報に対しても適用可能であること。

我々はユーザの興味を計算するメカニズムとして、ユーザの興味と対象情報の双方をそれぞれの特徴を表すベクトルで表現し、両者間の距離により、類似性を判断する方式として vector space model [5] を採用した。これにより、ユーザによる事前の定義が不要でかつ暗黙の学習が可能である。しかし、同モデルでは全情報を一様に管理するため、複数の観点からの興味を分類して管理できない、大量のキーワードに対して効率よく検索できないと言う欠点がある。

分野依存でかつ効率的な学習のためには、分割された記憶を切り替えて使用する必要がある。記憶の分割方式として、(1) 予め分類を定義し、その分類の中でのみ学習を行う、(2) 予め、分類は定義せず学習に従って興味の分類を動的に構成する方式、の2通りの方法が考えられる。前者は、定義が固定的で情報の変化に柔軟に対応できず、後者は、学習に膨大なサンプルが

では、Bさんに釣りの情報についても高い評価が与えられる。」²例えば、以下のような理由付けがユーザには有効であると思われる。(1)「あなたが好きな artist X を好きな user Yさんは、artist Zも気に入っている」(2)「あなたが好きなジャンル XではYが一番人気がある」。(1)については、従来方法でも可能、(2)については、対象の中身の情報を用いる必要がある。

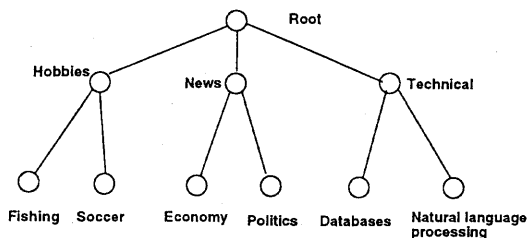


図 1: 記憶の静的構造の例

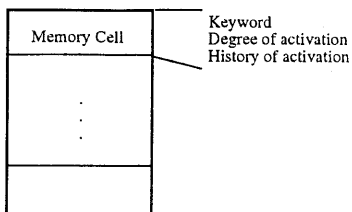


図 2: 記憶ノードの構造

必要という問題がある。我々は予め設定した分類を学習の過程で融合することにより、分野別の興味に沿った分割が可能な階層記憶方式を考案した。

次の章で、この階層記憶モデルを定義する。

3 記憶のモデル

3.1 記憶の静的構造

ユーザの興味を分類して保持するために、分野に従って名前付けられたノードを要素とする木構造を定義する。これは汎用的、かつ静的な定義であり、分野名を表す辞書に相当する。その例を図 1 に示す。

各々のノードは、その分類に含まれる記憶セルの集合を持つ (図 2 参照)。個々の記憶セルは、キーワード、そのキーワードの活性度、活性履歴からなる。キーワードの活性度 L は、ユーザのキーワードの興味の度合を表すもので、 $-1 \leq L \leq 1$ の範囲の値をとる。活性度はユーザの操作により記憶に対する刺激が発生することで更新され、値が高い程興味の度合いが高いことを示す (詳細は 3.3 で述べる)。また活性履歴は、学習の過程で活性化されたキーワードの活性度およびその日時を表され、忘却処理で用いられる (詳細は 3.3.2 で述べる)。

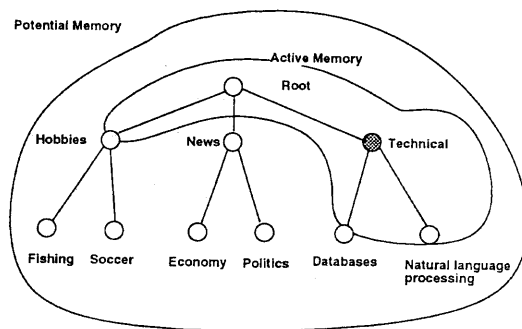


図 3: 顕在記憶

3.2 記憶の動的構造

興味の計算に記憶を適用するには、まず、全体の記憶構造 (潜在記憶と呼ぶ) から、今対象としている分野に対する記憶構造 (顕在記憶と呼ぶ) を動的に生成する。顕在記憶は、ユーザが現在対象としている記憶ノード (対象記憶ノードと呼ぶ) を中心に関係する記憶セルを集めることによって実現される。図 3 に、対象記憶ノード Technical に対して顕在記憶を生成した例を示す。

続いて、潜在記憶から顕在記憶を生成する手順を順に述べる。

3.2.1 対象記憶ノードの決定

対象記憶ノードは、ユーザが目している潜在記憶中のノード (分野) を表す。これは、複数のノードを指定することもできる。対象記憶ノードが複数ある場合は、それぞれのノードに対して 0 以上 1 以下の値の重み付けがなされる W_n 。

3.2.2 ノードの関連度の計算

顕在記憶を構成するには、まず、対象記憶ノードをその近隣の記憶ノードとの関連度 $R (0 \leq R \leq 1)$ を決定する。その手順を以下に示す。

距離 1 のノードの関連度 対象記憶ノードに含まれる記憶セルの活性度の絶対値の総和を T 、キーワードが近隣ノード中に含まれ、かつ、活性度の符号が等しい対象記憶ノードの記憶セルの活性度の絶対値の総和を C とすると、距離 1 のノードの関連度 R は以下の式で表される。

$$R = \frac{C}{T}$$

距離 $N (> 1)$ のノードの関連度 まず、対象記憶ノードとの関連度を求め、対象記憶ノードとの間のノードの関連度をすべて掛け合わせる。ただし、あるノードと対象記憶ノードの関連度がある一定の値より低くなった場合は、それ以上先のノードに対する関連度は 0 とする。

このようにして、対象ノードと他のすべてのノード間の関連性が計算される。

3.2.3 顕在記憶の生成

次に、関連性が 0 より大きなノードにおいて、すべての記憶セルの活性化にノードの関連度を掛け合わせる。対象記憶ノードが複数の場合は、それぞれのノードにおいて、さらにそのノードの重みが掛けられる。

それを活性化の大きな順に並べ、先頭から、あらかじめ決められた顕在記憶のサイズ個の記憶セルを集め、これを顕在記憶とする。

以上のようにして生成された顕在記憶は、注目している分野における興味の度合を計算するために用いられるが、その利用方法は実現する機能に応じて異なる。情報検索における機能に対する具体的な興味の計算方法は後述する。

3.3 記憶の活性化—学習と忘却

記憶は、外からの刺激と時間経過によって活性化される (学習と忘却)。刺激は、キーワードとその活性化との対の集合によって表される (情報検索における刺激については、後述する。)

記憶の活性化は、ユーザが 1 つの分野に対して連続した作業を行なっている間 (セッションと呼ぶ) は、顕在記憶に対して行なわれる。セッション終了時に、そのセッション中に生じた刺激を潜在記憶に適用した後、さらに、忘却の処理によって潜在記憶の活性化を行なう。

3.3.1 学習

刺激による顕在記憶および潜在記憶の活性化の方法を以下に示す。

刺激による顕在記憶の活性化 刺激の個々のキーワードの活性化を顕在記憶中の記憶セルに加える。同じキーワードを持つ記憶セルが存在した場合は、1 つの記憶セルにマージする。その記憶セルの活性化は、刺激の活性化と存在した記憶セルの活性化を加える。活性化が 1 より大きくなる、あるいは、-1 より小さくなる場合は、それぞれ 1、-1 とする。

刺激による潜在記憶の活性化 対象記憶ノードを始めに、個々の刺激のキーワードの活性化と、それに対するノードの関連度を掛けた値をそのノードの記憶セル

に加える。同じキーワードを持つ記憶セルが存在した場合は、1 つの記憶セルにマージする。その記憶セルの活性化は、刺激の活性化と存在した記憶セルの活性化を加えたものとする。活性化が 1 より大きくなる、あるいは、-1 より小さくなる場合は、それぞれ 1、-1 とする。

これらの動作を関連度が 0 より大きくなるすべてのノードで計算する。対象記憶ノードが複数ある場合は、同様の処理を繰り返す。

3.3.2 忘却

全体の記憶の利用効率を高め、ユーザの興味の時間的な変化に対応するために、記憶は時間経過によっても活性化される。この活性化は、用いられなかった記憶セルの活性化を下げることによって実現される (忘却)。これは、前述の「刺激による記憶の活性化」においてマイナスの刺激を伝搬させる場合と同様の処理による。

マイナスの刺激を生成するには、個々の記憶ノードに対して以下の操作を行なう。

- システム終了時に、そのセッション中活性化が上がらなかったすべての記憶セルの活性化を K ($0 < K < 1$) 倍する。
- キーワードの最終活性化日時によって、その活性化を期間に応じて減少させる。
- 一定期間に一定回数以上の活性化がなされた記憶セルは、活性化を下げる度合を減らす。

4 個人記憶モデルの情報検索への適用

前述した個人記憶モデルを情報検索に適用すると、記憶モデルの構成要素は以下のように対応する。

潜在記憶 文書の分野の階層構造

対象記憶ノード 文書の分野の階層構造中で、ユーザが今検索の対象としている分野。これは、対象とするデータベースの情報から自動的に決定できる。例えば、図 1 の例では、対象が論文データベースならば Technical に決定することができる。

顕在記憶 ユーザが今検索の対象としている分野を対象としたユーザ個人の興味の情報

刺激 文書の通読傾向、および、ユーザの指示

次に、情報検索において、記憶に対する刺激を生成するための手順について述べる。

4.1 ユーザ動作とその対象の認定

ユーザの興味をできるだけユーザに意識させずに学習するために、以下のようなユーザの動作から記憶に対する刺激を生成する。

入力されたキーワード 検索のためにユーザが入力したキーワード

文書の内容の表示 ユーザの動作によって画面上に表示された文書の内容

文書の保存 ユーザの動作によって保存された文書の内容

これらの動作は、ユーザの検索行動から自動的に収集することができる。しかし、ユーザの検索行動だけでは、興味のプラスの情報しか取得することができないし、明らかに好ましいことがらを陽に指定することもできない。そのために、ユーザに対し文書、あるいは語句ごとにそれらが、「重要」「不要」の判断をすることによってこれらの情報を獲得することも可能である。これらの動作に対しては、-1以上1以下の重みが付けられる。

4.2 ユーザ動作からの記憶に対する刺激の生成

次に、ユーザの動作の対象となるものから、その対象の中で重み付けを行なったキーワードを抽出し、それらに、個々の動作に対して付けられた重みを掛けたものを記憶に対する刺激とする。

対象の中での重み付けは、以下の処理による。

まず、入力語句を形態素解析を用いて品詞付の単語に分解する。それぞれの単語について、名詞でないものと名詞であるがキーワードとしての禁止語リストに入っているものを除外し、それ以外の名詞の単語をキーワードする。それぞれのキーワードについて以下のような特徴項目毎に点数を付け、それぞれの項目の重みと掛け合わせたものの和をそれぞれのキーワードの点数優先度)とする。

特徴項目としては以下のようなものが挙げられる。

- 出現回数：その単語のそのテキスト中での出現回数を点数とする。
- 助詞「は」：その単語の直後に助詞「は」がある場合に1点。
- 助詞「は」以外：その単語の直後に「は」以外の助詞がある場合に1点。
- カタカナ・アルファベット：その単語にカタカナ・アルファベットが含まれる場合に1点。

```
title:
近鉄ケーブルネット、CATVで関西初のゲーム配信。
content:
関西でも有線テレビ(CATV)の回線を通じてゲームソフトを配信するサービスが始まった。
奈良県生駒市と奈良市の一部を加入地域にするCATV会社、近鉄ケーブルネットワーク(KCN,奈良県生駒市)が始めたもので、サービスの名称は「セガチャンネル」。加入者はセガ・エンタープライゼスの家庭用ゲーム機「メガドライブ」と、受信用の専用レシーバークートリッジを用いて、常時三十種類のゲームを楽しむことができる。毎月五一六タイトルのゲームが入れ替わる。
```

図4: 新聞記事

```
ゲーム 1.000000
サービス 0.923077
生駒市 0.846154
奈良県生駒市 0.807692
奈良県 0.769231
加入者 0.576923
名称 0.576923
レシーバ 0.538462
ゲームソフト 0.538462
カートリッジ 0.538462
専用レシーバークートリッジ 0.487180
CATV会社 0.461538
...
```

図5: 抽出された刺激

- 重要文：その単語が重要文の中に含まれていれば、その文の数を点数とする。

それぞれの特徴項目の重みはヒューリスティクスにより以下のように設定できる。

```
出現回数 1
助詞「は」 0.5
助詞「は」以外 0.2
カタカナ・アルファベット 0.2
重要文 2
```

重要文は、接続詞、文末表現、文の位置などの表層上の手がかりからヒューリスティクスにより決定される。

このようにして、対象から、キーワードとその重みの対が決定される。この対に対して動作の重みによって重み付けし、記憶に対する刺激とする。

例として、入力記事(図4)から抽出された刺激を図5に示す。

4.3 刺激による記憶の活性化

このようにして得られた刺激は、前述した手法によって、潜在記憶の活性化のための入力となり、ユーザの

興味が蓄積される。

検索セッションにおける学習の手順は、以下のステップによる。

1. 検索セッション開始時に、決められた対象記憶ノードに対して顕在記憶を生成する
2. 同一検索セッション中に生じた刺激を顕在記憶に対して適用する。
3. 検索セッション終了時に、そのセッション中に生じた刺激を潜在記憶に適用する。さらに、忘却の処理を行なう。

5 個人記憶モデルを用いた検索支援機能

以上述べた階層的記憶を用いたユーザの興味に応じた検索のための機能として、文書のランキングと、キーワードネットにおける応用例について述べる。

5.1 ランキング

5.1.1 計算方法

ランキングとは、検索された文章をユーザの興味の順に並び換える機能を意味する。

階層的記憶構造を用いてランキングを行なうには、まず、ユーザの指定による分野の情報から対象記憶ノードを決定し、そこから顕在記憶を構成する。

次に、全ての(ランキングの対象となる文書集合および顕在記憶中の)キーワードを辞書順に並べて、 $1 \sim N$ (= キーワード数) の番号を付ける。文書集合と顕在記憶は共に、 $\mathbf{x} = [x_1, \dots, x_N]^T$,

$$x_i = \begin{cases} \text{キーワードの重み} & \text{if } i = \text{キーワードの番号} \\ 0 & \text{otherwise} \end{cases}$$

として、 R^N のベクトルとみなせる。各文書ベクトル \mathbf{d} と顕在記憶中のキーワードベクトル \mathbf{w} の内積

$$\mathbf{d} \cdot \mathbf{w} = \sum_{i=1}^N d_i \times w_i$$

を計算し、この値の降順で文書集合をソートする。

5.1.2 実行例

以下にキーワード“ゲーム”で検索された9記事を時間順、興味順に並べた例を示す。

時間順

- A ソフトビジネスの試練総合家電メーカーの苦悩 (下) 崩れる「効率優先」の神話
- B ゲームの疲れ軽減、カラーテレビ——船井電機 (ニューフェース)
- C マイクロキャビンソフト開発加速、CD-ROM ゲーム、年7—8本ベースに
- D 近鉄ケーブルネット、CATVで関西初のゲーム配信
- E 大阪に関西初のフリースクール、人との触れ合いを——ひきこもる若者自立支援
- F 吉本興業、上海でTV番組、若者向けバラエティー——現地局と共同制作
- G マルチメディア時代を模索するハリウッドビジネス——映画評論家菅見有弘氏 (月曜版)
- H “ペット飼える”電子手帳——芸や世話を疑似体験 (新製品比べてみると)
- I 大リーグ流で活——「トライハードだ！」練習もゲーム形式で (ヒトめぐり)

5.1.3 興味順

記事「“ペット飼える”電子手帳——芸や世話を疑似体験 (新製品比べてみると)」の内容を読んで学習させた後、興味順に並べた例を示す。

- H “ペット飼える”電子手帳——芸や世話を疑似体験 (新製品比べてみると)
- C マイクロキャビンソフト開発加速、CD-ROM ゲーム、年7—8本ベースに
- A ソフトビジネスの試練総合家電メーカーの苦悩 (下) 崩れる「効率優先」の神話
- E 大阪に関西初のフリースクール、人との触れ合いを——ひきこもる若者自立支援
- D 近鉄ケーブルネット、CATVで関西初のゲーム配信
- B ゲームの疲れ軽減、カラーテレビ——船井電機 (ニューフェース)
- I 大リーグ流で活——「トライハードだ！」練習もゲーム形式で (ヒトめぐり)
- G マルチメディア時代を模索するハリウッドビジネス——映画評論家菅見有弘氏 (月曜版)
- F 吉本興業、上海でTV番組、若者向けバラエティー——現地局と共同制作

5.2 関連キーワード表示機能

関連キーワード表示機能は、あらかじめ与えられているシソーラス、および、文書集合中の共起情報から計算された関連するキーワードを表示することによってユーザの検索要求に合うキーワードを想起するのを支援する機能である。

関連キーワード表示機能は、ユーザが指定した単語に関連する単語を表示する。この表示の際に、顕在記憶によって、活性度を計算し、その活性度順に表示することによって興味の順に関連するキーワードを見ることができる。

また、ユーザが、表示されたキーワードの中から自分の興味のあるキーワードを選択することによって、興味を学習できる。

6 個人記憶モデルの協調的フィルタリングへの適用

6.1 アプローチ

これまでに個人の興味を蓄積するための記憶モデルについて定義した。このモデルを協調的フィルタリングに適用することを考える。

個人記憶モデルにおいては、すでに「興味」が解析されており、それが階層的に管理されている。従来の協調的フィルタリングは個人を単位として計算を行っていたが、個人記憶モデルを導入することによってこれを記憶モデル中の個々のノードを単位として計算を行なう。これによって、従来の協調的フィルタリングと違い、個人の興味の多様性を取り扱うことが可能となる³。

また、ノードと情報との類似度が計算できるため、誰からもアクセスされなかった情報に対しても興味の度合いを計算することが可能である。さらに、ノード間の類似度が計算できるため、同じ情報を参照していないユーザ間の類似度を計算できる。

6.2 実現方法

個人の評価情報を個々の記憶ノード毎に管理することによって、前述の機能を実現することができる。利用の対象としては、情報の推薦と、評価依頼の2つが考えられる。

6.2.1 情報の推薦

推薦するための方法を以下に述べる。

1. 顕在記憶と他のユーザの記憶ノードのすべての類似度を計算する
2. 特に類似したユーザの記憶ノード群を決定する
3. それらの記憶ノード群で評価の高かった情報で推薦対象のユーザがまだ参照していないものを推薦する

6.2.2 情報の評価依頼

協調的フィルタリングにおいては、誰かによって評価がされなくては有効に機能しない。そのため「組織的な」評価の割り当てが有効となる。このために、以下の方法を用いる。

³例えば次のようなことが可能となる。「過去の評価の履歴から、Aさんは、「情報検索」関係に興味があり、Bさんは、「情報検索」関係と「釣り」関係に興味があるということをあらかじめ区別しておく。Aさんは、Bさんが高い評価を与えたものの中から、「情報検索」関係の情報だけを受けとる。」

1. 顕在記憶と他のユーザの記憶ノードの類似度を計算する
2. 特に類似したユーザの記憶ノード群を決定する
3. それらの記憶ノード群を用いてすべての情報をランキングし、評価の高かったもので、その記憶ノード群において参照されていない情報を「評価依頼」する。

この方法では、必ずしも個人の興味に一致していないが、集団行動として最適化される。

7 システムの実現

現在、今までに述べたモデルに基づくシステムを開発中である。データベースとして、日本経済新聞社の記事を用いている。システムの構成を図6に示す。

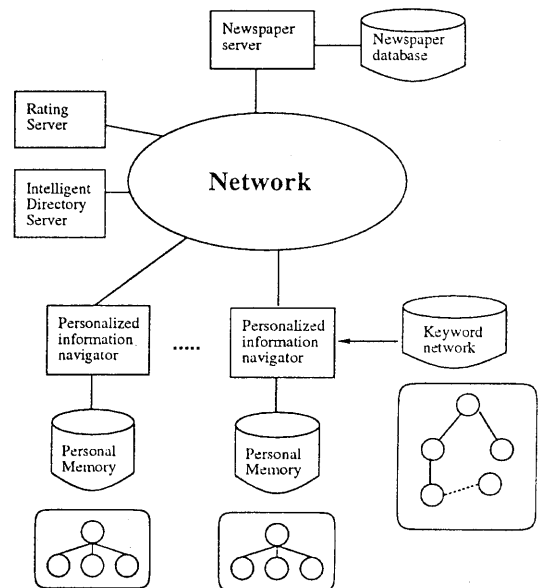


図6: システム構成

8 考察

8.1 個人記憶モデル

分野別の興味の管理 分野毎に記憶を構成し階層構造を持たせ管理することによって、今検索の対象としている分野の興味で検索支援を得ることが可能になった。さらに、分野間の関連度に応じてキーワードの活性度

を伝搬させることによって他の分野で興味のある事柄に関しても興味の度合が高ければ、それを反映することが可能である。

効率 学習を続けていくと個人に対して蓄積されるキーワードの数はかなりの数になり、検索における様々な処理が遅くなることが考えられる。特に、興味の度合を計算するプロセスは、対話的な環境で実行されることが予想されるため即座に回答を返すことが要求される。しかし、本手法は、実際の処理に用いる際に顕在記憶のサイズを小さく持つことによって、処理のスピードを落さずに行うことが可能となる。また、計算能力が十分にある、あるいは、計算時間が多く掛かっていても良い場合は、顕在記憶のサイズを大きくすることによってより細かい興味を反映できる。

顕在記憶の生成、および、記憶の活性化についても、蓄積されるキーワードの数が多くなるとその処理に時間が掛かるようになるが、これらの処理は、必ずしも対話的な環境での実行を要求されないため、検索終了時に計算する、あるいは、一定期間毎に計算することによってユーザの対話処理のスピードに影響しないようにすることが可能である。

評価 本システムでは、文書の自動分類などで用いられている vector-space model をユーザの観点で文書をランキングするのに応用している。文書に含まれるキーワードおよびその重みを用いて、各文書をベクトルで表現し、ユーザの興味をこのベクトル空間中のワードベクトルで表現している。現在のシステムでは、文書を参照した回数に基づいてワードベクトルの学習を行っている。本手法はユーザの興味を学習し、個人の観点による文書のランキングを目的としているため、適合率や再現率といった一般的な評価は適用できない。そこで、以下の手順によって本手法を評価することを考えている。

Step1 予め文書集合をユーザの観点によって興味のある文書群と興味のない文書群に分け、興味のある文書群に高いランキングを与える理想的なワードベクトルを作成する。

Step2 興味のある文書群から文書を参照することでワードベクトルを学習させ、理想的な場合と比較する。

2つのベクトルの類似度が高いほど良い学習が行なわれたことが分かる。学習量(参照した文書数)とベクトルの類似度の関係を評価することで学習速度が評価できる。

さらに、一過性の興味や操作の誤りによって参照された文書による副作用の大きさやユーザの時間的な興味の変化への対応、などを考慮する必要がある。

8.2 協調的フィルタリングへの適用

本稿では、通読傾向の類似性を興味の類似性の計算に考慮しなかった。これは、通読傾向が類似していれば、その結果生成される記憶も類似するという仮定に基づいている。しかし、通読傾向の類似性を考慮するような方法も考えられる。

また、推薦と評価依頼の2つの利用方法において、記憶モデルから類似ノード群を計算したが、あらかじめ決められたグループを定義することによってグループウェアとしても有効であると考えられる。

9 おわりに

本稿では、ユーザの興味に応じた情報検索を実現するため、ユーザに特別な操作を行なわせることなくユーザの興味を学習する手法、および、ユーザの複数の興味を効率良く扱えるような記憶構造のモデルを提案した。本モデルは、ユーザの興味を概念構造の部分集合で表現し、複数の分野の興味を統一的に扱うこと、および、情報への参照傾向からユーザの興味を獲得することを特徴とする。また、このモデルを協調的フィルタリングに適用する方法を提案した。

本モデルの有効性を評価するため、文書の参照回数によってユーザの興味を学習する機能、および、文書をユーザの興味の順に並べるランキング機能を持つプロトタイプシステムを試作した。今後は、ユーザの興味を学習する方法の評価、および、協調的フィルタリングシステムの実現を行なう予定である。

参考文献

- [1] Pattie Maes, "Agents that Reduce Work and Information Overload," Communications of ACM, July 1994, Vol.37, No.7
- [2] Upendra Shardanand, Pattie Maes, "Social Information Filtering: Algorithms for Automating Word of Mouth," CHI '95.
- [3] David Goldberg, David Nicholas, Brian M. Oki, and Douglass Terry, "Using Collaborative Filtering to Weave an Information Tapestry," Communication of ACM, Dec. 1992, Vol. 35, No. 12.
- [4] Paul Resnick, Neophytos Iacovou, et al., "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," Proc. of CSCW '94.
- [5] Salton G., "Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer," Addison Wesley, 1989.