

1 はじめに

WWW のような情報源からユーザが目的の情報を収集するのをサポートするエージェントが多数開発されている。WWW からの情報検索のように、特に個人の興味や嗜好がユーザの目的に大きな影響を与える場合、システムが個人に適応するためにユーザモデルの獲得が必要である。情報検索システムにおけるユーザモデルとは、ユーザの検索要求を特定するためのモデルであり、ユーザの嗜好と価値、期待、志向性、情報検索知識を表す [1]。

本研究では Web ブラウジング中のユーザが現在どんなドメインに興味を持っているのかをユーザモデリングによって推定し、エージェントが持っているドメインに特化したモデルとのマッチングを行う。マッチングの結果、適合したドメインに関する情報をユーザに提供することによりユーザの情報収集をサポートする。

2 ユーザモデリング

Belkin は情報検索システムにおけるユーザモデルは次の三点を満たす必要があると述べている [2]。(1) モデルはユーザの複数の興味を表現できなければならない。(2) モデルは時間によって推移するユーザの興味に追従できなければならない。(3) モデルはユーザの興味の変化をユーザの観察から得なければならない。本研究では (1) を満たすために、ユーザモデルを特徴ベクトルから構成される複数のクラスタによって表現した。(2) を満たすためにクラスタを構成する特徴ベクトルに対してユーザの興味と減衰を考慮に入れて重み付けを行った。また (3) を満たすために Web ページの内容に関する直接のフィードバックを必要としないことにした。本研究では Web ページに含まれるユーザの興味 (トピック) は一つであると仮定をする。また、本研究ではユーザは自分の興味に基づいて Web ブラウジングをしているとし、無目的な、つまり気まぐれなブラウジングは扱わないものとする。

以下にユーザモデリングの手順について述べる。

まず対象となる Web ページのソース (以下ドキュメント) から HTML タグを除去した本文に対して形態素解析を行い、名詞のみを取り出し出現頻度を求める。次にこの出現頻度にユーザの興味と減衰を反映した重み付けを行う。松田 [3] らは、ユーザに対して興味を与えずにブラウジングを行わせると、ユーザの興味は読んだノードに影響され推移し、ノードを読み進めるにしたがって過去に読んだノードが現在の興味に与える影響が減衰する傾向が見られることを、ブラウ

ジングタスクの分析によって確かめた。このことはブラウジング履歴を過去に遡っていくにつれて、現在見ている Web ページとの関連が薄れていくことを示唆している。本研究ではユーザの興味の対象がノードではなく Web ページの内容であるがこの傾向が適用できると思われる。 j 番目のページのキーワード i の出現頻度を tf_{ij} とすると、 tf_{ij} は以下の式で与えられる。

$$tf_{ij} = W \frac{\text{freq}(ij)}{\max_{ij} \text{freq}(ij)}$$

$$W = r^{n-j} (0 < r < 1)$$

ここで r はユーザの興味と減衰率を表すパラメータで、 n はユーザが今まで辿ってきたページの数である。これによりドキュメントの特徴ベクトルは m 次元のベクトル (m はページ中のキーワードの数) で表現される。

システム利用当初はユーザモデルが存在しないため、ある一定数特徴ベクトルが蓄積された後クラスタリングによりユーザモデルの初期化を行う。クラスタ間の類似度比較にはクラスタの重心を利用する。ユーザモデルの初期化が行われた後、新たな特徴ベクトルが追加された時、その特徴ベクトル一つを含むクラスタが生成され、各クラスタの重心との類似度が高いクラスタと融合される。

このようにして得られたクラスタの集合でユーザモデルを表現する。一つのクラスタがユーザの一つの興味を表現している。ここで、現在ユーザが見ているページを表す特徴ベクトルが含まれているクラスタを現時点でのユーザの興味とする。

3 ドメインモデルの構築

自分と興味と似ているユーザは自分と同じように情報を評価するという仮説に基づき、情報を提供するシステムを推薦システムという。推薦システムの中に Web ページを推薦するものがあるが、それらのシステムではドメインに特化した情報や知識を利用しないため、ユーザの興味に十分に対応できなかった。さらに、他のユーザが一度見た情報でないと、他のユーザには推薦されないという欠点があった。

本研究におけるドメインモデルは、以下の三つの要素から構成されている。(1) ドメインを表すキーワードの集合、(2) ドメインを特定するためのルール、(3) ドメインに特化した Web ページ (実際には URL) の集合。(1) および (2) はユーザモデルとの類似度を計算するために利用され、あらかじめドメインごとにカテゴリ分けされた Web ページの集合を用いて抽出される。

本研究ではドメインを表すキーワードの集合を抽出するために Yahoo! JAPAN のカテゴリを利用した。Yahoo! JAPAN の各ディレクトリは人手によってカテ

ゴリ分けされたものであり、そこに含まれる Web ページにはドメインの特定を妨げるノイズ、つまりドメインとは関係がないキーワードが少ないことが期待できる。プロトタイプシステムではこれらのディレクトリからコンピュータ、スポーツ、旅行関連を選び利用した。これらのページからドメインに特化したキーワードを多く抽出するために TFIDF を利用した。ここで、IDF 値を計算するための対象となるページは他のカテゴリに含まれるものとした。こうして抽出されたキーワードを重み付けし、重みが大きい方からドメインごとに 10 つのキーワードを選んだ。これが (1) のドメインを表すキーワードの集合となる。

さらにこれらのキーワードの組み合わせからドメインを特定するためのルールを構築する。例えば以下は現在のユーザの興味が Mac に関するカテゴリであると判断するためのルールである。

"Mac"&" ← Win"&" ← Unix" → "Mac"

"Apple"&"Computer" → "Mac"

以上のようなルールによってユーザモデルをドメインモデルに当てはめる。もしルールを適用してもドメインが特定しないときは、キーワードの集合から生成されるドメインの特徴ベクトルとユーザモデルに含まれるクラスタの重心との類似度を求めドメインを特定する。

また (3) のドメインに特化した Web ページの集合は Yahoo! JAPAN の各カテゴリからそのドメインの特徴ベクトルと最も類似したカテゴリ内のページを類似度の高い順に数ページ選ぶ。また、一般に公開されている検索エンジンを利用して新たにページを探し出し、検索エンジンが提供するスコアが高いページの URL を集合に追加することも可能である。このとき (1) のドメインを表すキーワードを検索エンジンに渡す。

4 WebRS

本研究ではユーザモデリングによって推定したユーザの興味に即した情報提供を行うシステム WebRS (Web page Recommend System) のプロトタイプを実装した。WebRS は MacOS8.5 以上で動作するアプリケーションとして実装されている。

4.1 システム構成

システムはブラウザコミュニケーションモジュールや HTML ソース解析モジュール、ユーザモデル構築モジュールなど、各機能ごとに独立した複数のモジュールによって構成されている。各モジュールの機能は互いに密着して連携しているため、各処理を直列に行ってはうまく動作しないことが予想される。例えば、説明モジュールは各モジュールの出力に応じてユーザに対する説明文を提示するので、あらかじめどのモジュールからの出力から説明文を生成するかというのは決めることができない。そこで、本システムでは各モジュールが黒板と呼ばれる共通のデータ構造を介して情報交換をする黒板モデルを採用した。各モジュールは独立した推論システムであり、個々のルールによって動作する。黒板は各モジュール共有のワーキングメモリと

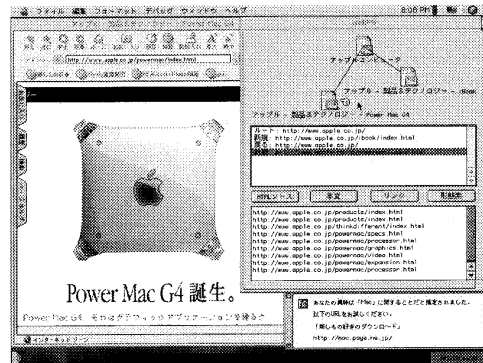


図 1: システム実行例

見なすことができる。各モジュールは黒板上に置かれたデータを自分のワーキングメモリに取り込み、照合・行動過程によって自らの行動を決定する。

4.2 システム実行例

図 1 に本システムの実行例を示す。図の左のウィンドウは Web ブラウザである。右上のウィンドウはユーザのブラウジング履歴と Web ページがどんな特徴ベクトルに変換されたのかをユーザが確認するためのウィンドウである。右下のウィンドウはシステムがユーザに URL を推薦するためのウィンドウである。

5 おわりに

本研究ではユーザモデルとドメインモデルのマッチングによってユーザの興味に即した情報を提供するシステムについて述べた。本システムではユーザモデルを特徴ベクトルで構成される複数のクラスタで表現することにより、ユーザモデルに対する要求を満たした。さらに、システムにドメインに特化した情報を持たせることにより、よりユーザの興味に即した情報を提供できるようなモデルを提案した。

今後の課題はシステムの実装を進め、本稿で述べたユーザモデルとドメインモデルのマッチングによるユーザへの情報提供が実際に有効なのかどうかを評価することである。

参考文献

- [1] Ingwersen, P., 藤原鎮男監訳, 情報検索研究: 認知的アプローチ, トッパン, pp.378, 1995.
- [2] Belkin, N., User modeling in Information Retrieval. Tutorial Overheads, available at <http://www.scils.rutgers.edu/belkin/um97oh/>, Sixth International Conference on User Modeling, Chia Laguna, Sardinia, 1997.
- [3] 松田憲幸, 野本豊裕, 平嶋宗, 豊田順一, 文脈情報に基づくブラウジング支援 - ブラウジング履歴の分析 -, 人工知能学会全国大会 (第 11 回), pp.456-459, 1997.