

## 利用者モデルの構想：情報フィルタによる利用者情報の収集

1M-3

下郡 信宏 月本 洋

株式会社 東芝 研究開発センター

システム・ソフトウェア生産技術研究所

## 1 はじめに

人間どうしの対話に比べ、人間とコンピュータの対話はどこちなく人間側に多くの負担を強いる。人間どうしの対話の場合、互いに相手のモデルを生成し、このモデルを用いてシミュレーションを行ないながら対話を行う為に除々に対話がなめらかになって行くと考えられる。これに対してコンピュータと対話する場合、人間はコンピュータのモデルを生成して行くが、コンピュータは何時まで経っても利用者に関するモデルを生成しない。これが、どこちない対話の一つの要因であると考えられる。本研究では、認知モデルに各利用者特有の情報を知識として持たせることにより、利用者モデルを構成することを提案する。

## 2 情報フィルタ

利用者モデルを作成するにあたり、その適用領域として情報フィルタを選んだ。情報フィルタ [1] とは、次々と到着する新しい情報（今回は文書情報）の中から利用者にとって有用であると思われる物だけを取り出すシステムの事である。利用者モデルを作成するには、利用者に関する膨大な情報が必要である。そこで、利用者が興味を持った情報を観察し、ここから利用者が興味を持っている領域に関する情報を取り出す方が類似検索等と比較しても情報量が多く効率的と思われる。また、情報フィルタならば日常的に接する為、より多くの情報が得られる事も期待出来る。

今回作成した情報フィルタは以下の手順で処理を行なう。

Modeling the User:

Gathering User Information from an Information Filter

Nobuhiro SHIMOGORI, Hiroshi TUKIMOTO

Systems & Software Engineering lab.,

R&D Center, Toshiba Corp.

1. 文書からキーワードを抽出
2. 文書の有用度を計算
3. 有用と判定されたものを利用者へ提示
4. 利用者からの有用度のフィードバックを取得
5. 利用者モデルの学習

## 3 利用者モデル

利用者モデルは人間を表現する仕組と特定の利用者を表現する知識から構成される。人間の認知活動をシミュレートするモデルとして認知モデルが幾つか提案されている。今回はそれら認知モデルの一つである ACT\*[2] を簡略化し、人間を表現する仕組として用い、これに各利用者特有の知識を付加して利用者モデルを得る。

ACT\* は宣言的知識（連想記憶）と手続き的知識（プロダクション）から構成されており、外界からの刺激により、宣言的知識上で活性伝播を行ない、活性化された宣言的知識上のノードの状態に最もふさわしいプロダクションを実行する事で、人間の認知活動をモデル化している。ACT\* の機能のうち、意識を制御するためのゴール構造は情報フィルタの性質上必要でない為、現在は実装していない。また、データ構造も Anderson の提唱する様な階層的なものではなく、キーワード間の継りを平面的に表すものである。学習機能は、ACT\* で提唱されている連想記憶の学習、プロダクションの強化、汎化、差別化などを実装している。

## 4 実験

作成した情報フィルタを用いて簡単な実験を行なった。被験者は筆者。フィルタリングの対象として利用した文書は社内のコンピュータ分野の特許提案書 2130 件。特許提案書には特許提案者、もしくは特許担当者により、適当なキーワードが 1~6 個付加されている。また、国際特許分類による分類コードも 1~3 個付加され

ている。

実験は以下の手順で行なった。

1. 点数付け 全 2130 件の文書に対して被験者が被験者にとっての重要性を 1～5 点の 5 段階で評価し有用である文書程高い点数を付けた。
  2. 分割 点数付けした文書をランダムに 10 等分し、1 セットを初期学習用、1 セットをテスト用、8 セットを学習用データとした。
  3. 初期学習 初期学習用データから 4 点と 5 点のものを用いて、初期学習を行なう。
  4. テスト 学習機構を止め、テスト用データを用いてフィルタリングを行なう。フィルタリングの結果、利用者に提示すると判断した文書を記録する。
  5. 学習 学習用データ 1～8 を用いてフィルタリングを行ない、利用者に提示すると判定した文書の評価値をフィードバックとして与え、学習を行なう。
- 上記のテスト、学習を全ての学習用データに関して繰り返し行ない、最後に再びテストを行なう。
- テストの結果は図 1 の通り。図はフィルタリングの結果、選ばれた文書の平均点と 4 点以上の文書の再生率を表している。

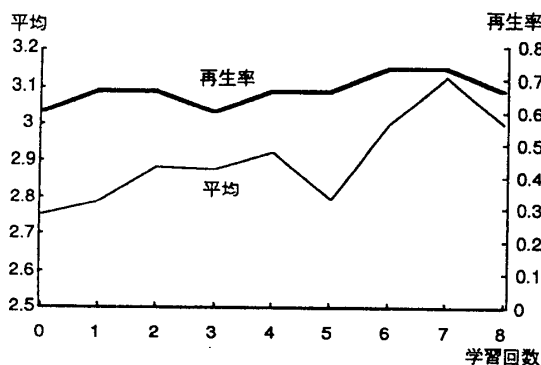


図 1: テスト結果

比較対象として、初期学習用データから、評価値が 4 点以上の文書の国際特許分類を取り出し、この分類に基づきテストデータの検索を行なった。その結果、テスト用データの平均点 1.94 に対し、国際特許分類を用いた検索で取り出した文書の平均点は 2.64 であった。

## 5 考察と今後の課題

フィルタリングにより選ばれた文書の平均点は国際特許分類による検索で得られた文書の平均点よりも良い値

が得られている。学習回数を重ねるにつれ、選ばれた文書の平均点が波打ちながらも徐々に上がっているのに対して、再生率はそれほど変化していない。これは、必要な文書を選び出す能力よりも、不必要な情報を捨てる能力の方が優れている為だと考えることが出来る。また、フィルタリングにより選ばれた文書の平均点が波打つのは、汎化学習によりプロダクションの適用範囲が広がり過ぎて不必要な文書を収集してしまった後に差別化学習によってプロダクションの適用範囲が狭げられる為に起こると考えられる。

今回の実験で観察された問題点とそれを解決する方針を示す。ACT\* の連想ネットワーク上で隣接するノードへの活性伝播は、隣接するノードの大きさに依存して決定される為、出現頻度の高い一般的なキーワード（例えば、データベースなど）は活性伝播の際に周辺ノードから多くの活性を奪ってしまい、周辺のノードの活性化を阻害する。これを解決する為に、活性伝播をノードの大きさ（ノードの出現頻度）により決定するのではなく、二つのノードの共出の頻度により決定する方法が良いと思われる。

今回用いた文書のキーワードは付加した人が文書毎に異なる為、バラバラな付け方がされている。例えば「文書検索」が一つのキーワードとして現れる一方、「文書」と「検索」の二つのキーワードに分けて付加されている場合もある。この様な違いは連想記憶の学習が進むと、次第に吸収されて来る筈だが、今回のデータ量では不十分であったと考えられる。将来は文書の本文から自動的にキーワードを抽出する予定である為、この様な間違いは起こらなくなると考えられる。

今後は情報フィルタで生成した利用者モデルを他のアプリケーションから利用する方法を検討する。

## 参考文献

- [1] Peter W. Foltz and Susan T. Dumais. Personalized information delivery: An analysis of information filtering methods. *CACM*, 35(12):51-60, 12 1992.
- [2] John R. Anderson. *The Architecture of Cognition*. Harvard University Press Cambridge, Mass., 1983.