

## 記事情報の選別フィルタリングにおける 興味の変化への追従性の改善に関する研究

山根 稔弘<sup>†</sup> 稲元 勉<sup>‡</sup> 樋上 喜信<sup>‡</sup> 小林 真也<sup>‡</sup>

<sup>†</sup>愛媛大学 工学部 情報工学科 <sup>‡</sup>大学院 理工学研究科

### 1. はじめに

近年、インターネットの活用機会が増加し、膨大な情報から必要な情報を選別することが困難である、情報技術を扱えるか否かの格差が生じる等の問題が生じている。我々この問題の解決を目的とした個人向け情報配信システム PINOT (Personalized Information On Television Screen)<sup>[1]</sup>の開発を進めている。

PINOT はインターネットを介して自動的に記事情報を受信し、ユーザの興味に基づき有用な記事情報をテレビ画面に表示する。表示された記事情報に対してリモコン操作が可能であり、操作履歴から記事情報に対する興味を類推する。そのため、不要と判別され表示されない記事情報には興味の類推が行われず、その記事情報に含まれる単語に対する興味の変化への追従性が低いという問題がある。そこで、本研究では各単語に対する興味の度合いに有効期限(寿命)を設定することで、各単語に対するユーザの興味の変化への追従性の向上を目指す。

### 2. PINOT

PINOT<sup>[1]</sup> は情報配信サーバ、情報 BOX, テレビ, リモコンで構成される。情報配信サーバはインターネットを通じて情報 BOX に記事情報を配信する。情報 BOX は配信された記事情報の選別を行い、テレビ局から放送される映像と選別した記事情報を合成し、右から左へ流れる(Ticker 形式)ようにテレビ画面の下部に表示する。ユーザは表示された記事情報に対しリモコンを用いた複数の簡単な操作を行うことができる。

PINOT の動作の流れについて説明する。

- ① 配信された記事情報を情報 BOX で取得する。
- ② 情報 BOX において、配信された記事情報に対し、ユーザプロファイルを参照し興味の度合いを計算する。ユーザプロファイルとは、類推された各単語に対するユーザの興味の度合いを学習し、記録しているものである。
- ③ 計算された記事情報に対する興味の度合いが閾値以上であれば、記事情報をテレビ画面の下隅に表示する。閾値未満であれば①に戻る。
- ④ 表示された記事情報に対するユーザのリモコン操作から、その記事情報に対するユーザの興味の有無を類推する。
- ⑤ 興味の類推結果をもとにユーザプロファイルを更新し、①に戻る。

②の興味の度合いの計算方法について手順を説明する。まず記事情報  $W$  に対して形態素解析を行い、単語を抽出する。抽出した各単語  $w_n (1 \leq n \leq N)$  に対し、ユーザプロファイルを参照して各単語の興味の度合い  $i(w_n) (0 \leq i(w_n) \leq 1)$  を取得する。ただし、 $N$  は抽出した単語の総数である。なお、抽出した単語がユーザプロファイルに含まれていない新出単語の場合は  $i(w_n) = 1$  とする。(1)式より記事情報  $W$  に対する興味の度合い  $I(W)$  を算出する。その値  $I(W)$  が閾値  $D (0 \leq D \leq 1)$  以上であればテレビ画面に表示し、閾値未満ならば表示しない。

$$I(W) = \frac{\sum_{n=1}^N i(w_n)}{N} \quad (1)$$

表示した記事情報に対するリモコン操作による興味の類推結果から、各単語  $w_n (n = 1, 2, \dots, N)$  に対し(2)式を適用してユーザプロファイルに格納されている単語の興味の度合い  $i(w_n)$  を更新する。

$$i(w_n) := \alpha \times i(w_n) + (1 - \alpha) \times J \quad (2)$$

ここで、 $J$  は表示された記事情報に対するユーザの興味の有無であり、「興味あり」なら 1 を、「興味無し」なら 0 をとる。また、 $\alpha$  は更新を行う際に古い興味の度合いと新しい興味の度合いが影響する割合を表す。

Improvement of adaptive ability of information filtering for articles according to the change of user's interest

<sup>†</sup>Toshihiro YAMANE, Tsutomu INAMOTO,

Yoshinobu HIGAMI, Shin-ya KOBAYASHI

Department of Computer Science, Faculty of Engineering, Ehime University (<sup>†</sup>)

Graduate School of Science and Engineering, Ehime University (<sup>‡</sup>)

### 3. 提案手法

本研究では、各単語に対する興味の度合いに有効期限を設定する手法を提案する。ユーザプロフィールに保存された単語全てに有効期限を付与し、一定期間更新されず期限切れとなった単語の興味の度合いを規定値  $\theta$  に変更する。興味が無いと判別され更新されにくくなった古い興味を一新し、素早く最新の興味を反映させることを狙う。

提案法では、各単語の有効期限を示すカウンタを用意している。記事に対してリモコン操作が行われるたびに、記事に含まれていない単語のカウンタを減少させ、カウンタが 0 となった時点で、その単語の興味の度合いを  $\theta$  に変更する。これによって興味の度合いが高くなり、その単語を含む記事が表示されやすくなる。

### 4. 実験

被験者を 3 人とし、複数の Web サイト上で約 8 カ月間配信されたニュース記事 2000 件を用いた実験を行った。本実験は興味の变化への追従性と情報フィルタリング精度を対象とした実験のため、リモコン操作による興味の有無の類推を行わず、全記事に対するユーザの興味の有無をユーザから直接得て使用した。

全 2000 件の記事に対し従来法と提案法の両方で情報フィルタリングを行い、その結果を以下の二つの評価基準を用いて評価した<sup>[1]</sup>。

- ① 情報 S/N 比：ユーザにとって興味があると判別し表示した記事情報の中におけるユーザが実際に興味を持った記事情報の割合のこと。計算式を(3)式に示す。ただし、ユーザにとって興味があると判別された記事情報の件数を  $H$  とし、 $H$  の中でユーザが実際に興味を持った記事情報の件数を  $U$  とする。

$$\text{情報 S/N 比(\%)} = \frac{U}{H} \times 100 \quad (3)$$

- ② 一致率：情報フィルタリングを行った全記事情報の中で、記事情報に対する興味の判別結果と実際のユーザの興味の有無が一致した記事情報の割合のこと。計算式を(4)式に示す。ただし、興味の判別結果と実際のユーザの興味が一致した件数を  $K$  とする。

$$\text{一致率(\%)} = \frac{K}{\text{全情報件数}} \times 100 \quad (4)$$

本実験では、実際に被験者の興味が変わるには時間がかかることから、使用する被験者の興味

を別の被験者のものに変更することで興味が変わったものと扱った。被験者は 3 人であるため、全 6 通りの興味の变化を用意した。また、閾値  $D$  と(2)式における  $\alpha$  の値を 0.6 とし、寿命のカウントの初期値として 100, 200, 300 を、 $\theta$  の値として 0.5, 0.6, 0.7 を調査した。

実験結果を評価するために、記事を 1 件フィルタリングするごとにその時点における情報 S/N 比と一致率を記録してグラフ化した。

#### 4.1. 実験結果

興味の各変化パターンにおいて、寿命のカウントの初期値を 300、 $\theta$  を 0.5 にすることでより高い効果が得られた。興味の变化のパターンのうちの一つの実験結果を図 1 に示す。実験結果より、情報 S/N 比、一致率共に大きな違いは確認できなかった。より多くの記事情報を取得するべく複数の情報サイトから記事情報を取得したため、類似の記事情報の比率が高く、記事のバリエーションが少なかった。このことが原因で評価に用いる記事系列として適切と言えなかったことが考えられる。

### 5. まとめ

本研究では、個人向け情報配信システム PINOT において興味の变化への追従性を向上させるための手法を提案した。提案法では、ユーザの興味を保存したユーザプロフィールにおける各単語に対する興味の度合いに有効期限(寿命)を設定した。本稿の実験結果からは、従来法と比較して興味の变化への追従性が向上したとは確認できなかった。

今後は、記事情報を一つの情報サイトから取得するなどして、より現実的な条件下で再度実験を行っていく予定である。

#### 参考文献

[1] 大岡 哲也. "Information filtering method using diversity among languages for personalized information delivery systems.", Polish Journal of Environmental Studies, Selected Paper of ACS 2009, vol.18, no. 4A, pp.67-71, 2009.

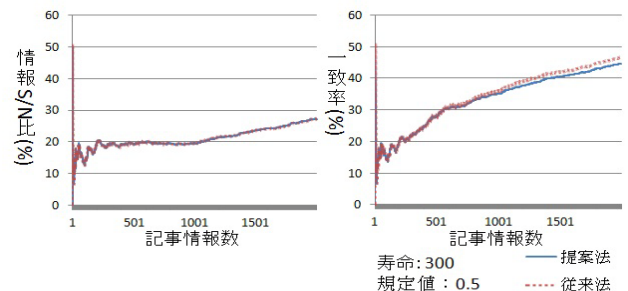


図 1 実験結果：情報 S/N 比(左)と一致率(右)