

個人向け情報配信システムにおける情報フィルタリング精度改善手法

大岡 哲也[†] 松村 和紀[†] 平田 孝志[†] 樋上 喜信[†] 小林 真也[†][†]愛媛大学大学院理工学研究科

1. はじめに

近年の情報通信機器及びインターネットの普及に伴い発生した情報過多やデジタルデバインド¹⁾の問題を解決するために、我々は個人向け情報配信システム PINOT (Personalized Information On Television screen) を構築している²⁾。PINOT では、サーバから送られてきた文字情報を、ユーザの興味の有無に応じてフィルタリングを行い、有用な情報のみをユーザに提供することを目的とする。このフィルタリングでは、文字情報を単語に分解し、それぞれの単語の興味の度合い及び、閾値に基づき、情報の表示・非表示を決定する。本稿では、これまでは別の単語として扱われていた類義語を同一の単語として扱うことで更なるフィルタリングの高性能化を行う手法を提案し、その有効性を示す。

2. PINOT

2.1. システム概要

PINOT は、情報を配信する情報配信サーバ、サーバとの情報の送受信やテレビへの情報の出力などを行う情報 BOX、テレビとリモコンで構成されている (図 1)。PINOT では、情報 BOX においてサーバから送られてくる情報を、情報に対するユーザの興味の度合いに応じて自動的に選別し、有用と判断された情報のみをテレビに表示する。また、テレビとリモコンによるインターフェイスを用いているので、ユーザに負担を掛けることなく必要な情報のみをユーザに提供することができる。すなわち、PINOT によって情報過多および情報リテラシーの有無によるデジタルデバインドの問題解消が期待できる。

2.2. PINOT のフィルタリング手法

図 2 に示すように、PINOT の情報フィルタリングは大きく分けて、「興味の度合いの計算」、「興味の有無の類推」、「ユーザプロフィールの更新」という 3 つの過程から成り立っている。また、ユーザプロフィールには、情報を構成する単語 w の興味の度合い $i(w)$ ($0 \leq i(w) \leq 1$) が格納されて

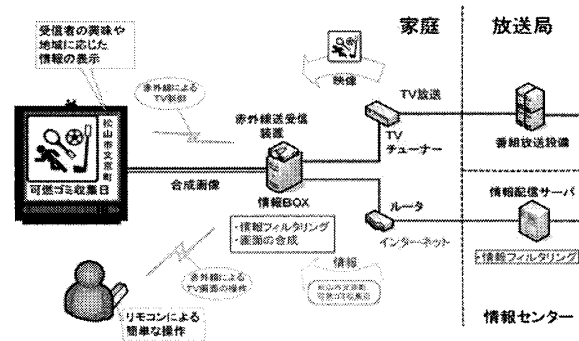


図 1: PINOT の概要

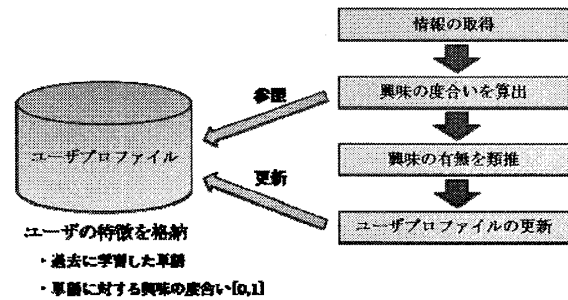


図 2: 情報フィルタリングの流れ

いる。まず、形態素解析ソフト茶筌³⁾を用いて配信された情報 W を構成する単語を抽出する。次に、抽出された単語 w_n ($n=1,2,\dots,N$) の興味の度合い $i(w_n)$ をユーザプロフィールから参照する。これを基に、配信された情報 W に対する興味の度合い $I(W)$ ($0 \leq I(W) \leq 1$) を以下の (1) 式で求める。

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

この興味の度合い $I(W)$ が閾値 θ 以上であれば、その情報を表示し、そうでなければ非表示となる。このとき、表示情報に対してユーザが行ったリモコン操作を基に、情報に対する興味の有無を類推し、(2)式を用いてユーザプロフィールにある各単語の興味の度合いを更新する。

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

ただし、 α ($0 \leq \alpha \leq 1$) はパラメータ、 J は表示された情報に対するユーザの興味の有無であり、「興味がある」なら 1 を、「興味が無い」なら 0 の値をとる。

An effective information filtering method for personalized information delivery systems

[†]The authors are with Graduate School of Science and Engineering, Ehime University.

3. 提案法

3.1. 提案法の概要

単語の中には、似た意味を持つ類義語が存在する。従来の PINOT では、単語とその類義語は、全く別の単語としてユーザプロフィールに格納されている。しかし、ユーザにとって単語とその類義語は同等の興味があると考えられる。そこで、本稿では、これら類義語を同じ単語として扱うことで、フィルタリング性能を向上させる手法を提案する。しかし、これを実現するためには類義語を効果的に判別しなくてはならない。提案法では、類義語の判別方法として、異なる言語間の変換の特性を利用する。つまり、ある日本語の単語を異なる言語の単語に変換した場合類義語は同じ単語に変換される可能性がある。その特性を利用して、提案法は具体的には以下の動作を行う。

まず、送信されてきた情報抽出した日本語の単語を、和英辞書⁴⁾を参照し英単語に変換する。変換後は、従来通りの手法でフィルタリングを行い、ユーザプロフィールに格納する際も英単語として格納する。英単語に変換できない場合は、変換せずに日本語のまま扱う。いくつかの類義語が一つの単語にまとまることで、各単語の興味の度合いの平均更新回数が増え、よりユーザの興味に沿ったユーザプロフィールの構築が可能となるのが期待できる。

3.2. 評価実験

提案法の有効性を示すために、フィルタリング手法の実証実験を行う。なお、被験者は 1 名である。また今回は、情報フィルタリング精度の評価のみを目的とするため、興味の有無の類推が 100% 正しい状態での評価を行う。そこで、文字情報として被験者に予め興味の有無を判別してもらった 3000 件のニュース情報を用いてフィルタリングを行う。パラメータ α 及び θ はそれぞれ 0.6 に設定する。ここで、評価尺度として以下の二つの指標を導入する。

$$\text{情報S/N比(\%)} = \frac{\text{表示された興味のある情報件数}}{\text{表示された情報件数}}$$

$$\text{再現率(\%)} = \frac{\text{表示された興味のある情報件数}}{\text{興味のある全情報件数}}$$

表 1 は、従来法及び提案法の情報 S/N 比及び再現率の平均値である。また、図 3 は情報 100 件毎の情報 S/N 比の推移を、図 4 は情報 100 件毎の再現率の推移を表している。これらの結果より、従来法に比べて提案法は情報 S/N 比及び再現率が全体的に高いことが分かる。よって、類義語を考慮することにより効果的なフィルタリングが行えるといえる。

表 1: 情報 S/N 比及び再現率の平均

	従来法	提案法
情報 S/N 比 (%)	69.9	73.5
再現率 (%)	93.5	96.2

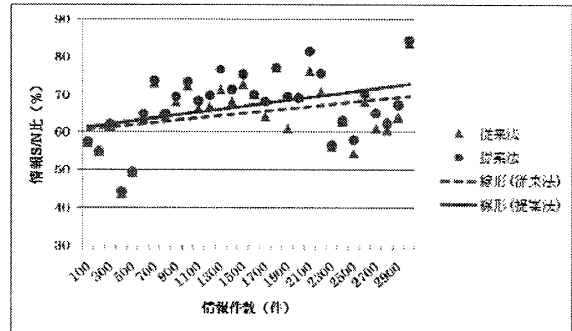


図 3: 情報 S/N 比の推移

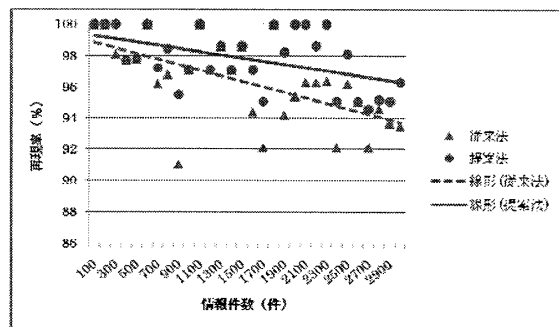


図 4: 再現率の推移

4. まとめ

本稿では、個人向け情報配信システム PINOT において、類義語を考慮した情報フィルタリング手法を提案した。提案法により、類義語を判断し、情報フィルタリングの性能を向上させた。

謝辞 本研究は戦略的情報通信研究開発推進制度 (SCOPE) 委託研究課題 No. 082309002 “TV を活用したプッシュ型地域情報配信システムの研究開発” の支援を受けている。ここに記して謝意を表す。

参考文献

- 総務省情報通信統計データベース, <http://www.johotsusintokei.soumu.go.jp/statistics/houdou05.html/>
- 小森 健市, 西岡 豊, 柏木 紘一, 樋上 喜信, 小林 真也, “個人向け情報の配信を目的とした PUSH 型情報配信システムの構築,” DICOM02005 シンポジウム論文集, July 2005, pp. 249-252.
- 形態素解析システム茶筌, <http://chasen.naist.jp/hiki/ChaSen/>
- 鍋田辞書, <http://www.nabeta.tk/>