

個人向け情報配信システムにおける 文字情報のフィルタリングに関する研究

遠藤 洋記[†] 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、有用な情報のみユーザに提供する。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。しかし、従来のフィルタリング方式では、本来興味のある情報が非表示にされてしまう可能性が高いという問題がある。そこで本稿では、ユーザが興味のある情報を取り逃さないような情報の選択を行うために、単語の興味の度合い及び閾値の関係を考慮した手法を提案し、その有効性を示す。

2. 個人向け情報配信システム Pinot

2.1. Pinot の概要

Pinot は、家庭外にテレビ映像を配信する放送

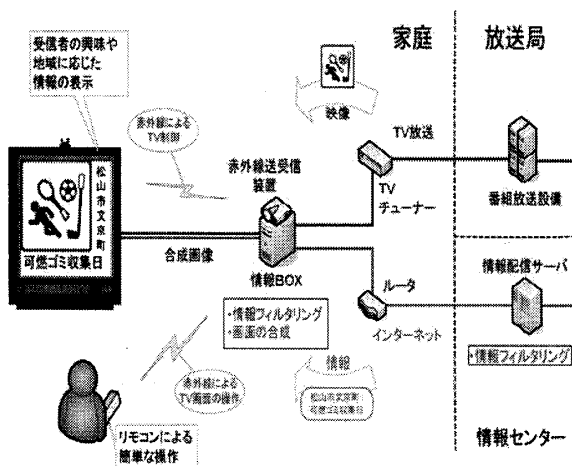


図 1 : Pinot のシステム

A filtering of text information for Personalized Information delivery System

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

表 1 : ユーザプロファイルの例

単語	興味の度合い $i(w)$
松山市	1.0
中島町	0.5
みかん	0.4
出荷	0.1
...	

局、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 とリモコン、テレビで構成されている (図 1)。

Pinot では、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に選別し、有用と判断された情報のみを表示する。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、ユーザはパソコンの様な情報通信機器を操作することなく、有用な情報を取得できる。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

Pinot では、ユーザプロファイルを用いて情報配信サーバから配信された文字情報を選別する。ユーザプロファイルは、単語 w (形態素解析⁽³⁾) によって得られた名詞と動詞) と単語に対する興味の度合い $i(w)$ ($0 \leq i(w) \leq 1$) で構成される (表 1)。このユーザプロファイルを基に、サーバから送られてきた情報に含まれている各単語に対する興味の度合いの平均を算出し、その値が閾値 θ を超えるとその情報を表示する。このとき、表示された情報に対してユーザが行ったリモコン操作の履歴を基に、情報に対する興味の有無を判断し、次式を用いてユーザプロファイルを更新する。

$$i(w) := \alpha i(w) + (1 - \alpha) J \quad (1)$$

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。

しかし、この手法では、頻繁に出現する単語の興味の度合い $i(w)$ が非常に下がりやすくなるという問題が存在することがわかった。つまり、一部の単語の興味の度合いが下がることで、「本来興味のある情報が非表示にされてしまう」可能性が高くなる。また、一度興味の度合いが低下してしまうと、その後上がりにくくなってしまふ。そこで、本稿では閾値 θ と興味の度合い $i(w)$ の関係を考慮し、興味の度合いの取りうる範囲を制限してやることでこの問題を解決する手法を提案する。

3. 提案手法

3.1. 提案手法の概要

提案手法では、既存手法で定義されていた興味の度合い $i(w)$ ($0 \leq i(w) \leq 1$) の範囲を $\delta \leq i(w) \leq 1$ とする。ただし $0 < \delta < 1$ とする。つまり、(1)式を用いてユーザプロファイルを更新する際に、下限値を δ と設定してやることで、無駄な興味の度合いの低下を防止する。ただし、(1)式により計算された $i(w)$ が δ を下回った場合は $i(w) = \delta$ とおく。これにより、無駄に興味の度合いが下がることなく、興味のある情報を効果的に拾うことができると考えられる。

3.2. 提案手法の評価

ここで、実際にユーザに興味の有無を入力してもらったデータを用いて 3000 件の情報にフィルタリングを行い、提案手法の評価を行う。ここで、閾値 $\theta = 0.6$ 、(1)式におけるパラメータ $\alpha = 0.6$ と設定する。

表2及び3はそれぞれ異なるユーザにおいて、 δ に対する SN 比及び再現率を表している。ここで SN 比とは (表示された興味ありの情報の数) / (表示された全情報の数) を示す。また、再現率とは (表示された興味ありの情報の数) / (全情報中で実際に興味がある情報の数) である。ただし、 δ が 0 の場合は既存方式を示す。これらの表より、 δ を増加させることで、再現率が効果的に上昇することがわかる。つまり、ユーザが必要としている「興味のある情報」を取り逃す可能性が低下している。また SN 比はほとんど変化がなく、効果的なフィルタリングが行えているといえる。

表2 : SN 比と再現率 (ユーザ A)

δ	0	0.2	0.3	0.4	0.5
SN 比 (%)	15.0	14.9	15.0	14.9	13.9
再現率 (%)	76.3	80.8	83.4	87.9	94.1

表3 : SN 比と再現率 (ユーザ B)

δ	0	0.2	0.3	0.4	0.5
SN 比 (%)	30.8	30.5	30.2	29.8	29.3
再現率 (%)	84.2	86.0	88.4	89.7	96.3

4. まとめ

本稿では、個人向け情報配信システム Pinot で使用されるフィルタリングにおいて、閾値と興味の度合いの関係に考慮したフィルタリング手法を提案した。提案手法により、SN 比を保持しつつ、従来手法よりも高い再現率を実現することができた。ただし、本稿においては固定の閾値しか用いていないため、今後は閾値も併せて最適に選択するような手法を考慮する必要がある。

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

- (1) 総務省情報通信統計データベース, <http://www.johotsusintokei.soumu.go.jp/statistics/houdou05.html>.
- (2) 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.
- (3) 形態素解析システム茶筌, <http://chasen.naist.jp/hiki/ChaSen/>