

操作履歴に基づき個人向けにニュースを選択表示する スマートフォンアプリの開発

Development of the Smartphone Application to Choose and Display News for Individuals based on an Operation History

小野 智士*
Satoshi Ono

稲元 勉†
Tsutomu Inamoto

樋上 喜信†
Yoshinobu Higami

小林 真也†
Shinya Kobayashi

1 研究背景

情報化社会の発展により、情報端末からインターネットを利用することで、容易に情報を得ることができるようになった。特に近年では、スマートフォンの普及が急速に進み、多くの人々がインターネットを利用する際の情報通信端末としてスマートフォンを使っている。

スマートフォン向けサービスの利用率の上位にニュース情報の取得が挙げられている。しかし、多くのニュースアプリは、配信されてきたニュースの全てをそのまま表示しただけである。そのため、ユーザは多くのニュースの中から興味のあるものを選別しなければならない。しかし、膨大な情報の中からの選別は、非常な時間と労力を要し、ユーザにとって大きな負担となる。このように、情報の量が膨大であるために、ユーザが有用な情報を選別することが困難になることを情報過多という。

本研究では、スマートフォンのニュースアプリにおける情報過多の問題を解決することを目的とし、個人向けにニュースを自動で選択し表示するニュースアプリ「スマートフォン版 PINOT」の開発を行う。



図 1: スマートフォン版 PINOT のシステム概要図

2 スマートフォン版 PINOT

スマートフォン版 PINOT とは、個人向け情報配信システム「PINOT」[1]の機能をスマートフォンでも利用出来るようにしたシステムである。スマートフォン版 PINOT システムは、スマートフォン上のアプリで構成される。スマートフォン上のアプリに対するユーザの操作から、ユーザがどのようなニュースに興味があるかを類推し、情報配信サーバから配信されたニュースを、ユーザの興味を考慮し表示する。また、情報配信サーバは、ニュースの記事情報をスマートフォンに配信する。

2.1 記事情報配信から表示までの処理

Step1 配信サーバがニュースの記事情報を配信。

Step2 アプリにおいて、記事見出し文を分割し、見出しの内容を特徴付ける名詞と動詞のみを抽出する。

Step3 過去の履歴に基づき、ユーザの興味に関する情報を記録したユーザプロファイルに基づき、抽出した各単語の興味の度合いを求め、その平均の記事見出し文の興味の度合いとする。

Step4 算出した興味の度合いが閾値以上の記事見出し文を表示。

Step5 ユーザの操作を受け付ける。

Step6 ユーザの操作履歴から、表示された記事見出し文に対する興味の類推を行う。

Step7 類推結果を基に、ユーザプロファイルを更新。

*愛媛大学大学院理工学研究科電子情報工学専攻情報工学コース

†愛媛大学大学院理工学研究科電子情報工学専攻

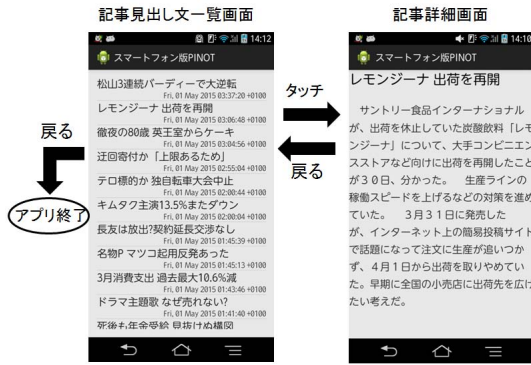


図 2: スマートフォンの表示画面

2.2 記事情報の表示形式

記事情報の表示形式は、多くのニュースアプリと同じく、記事見出し文から記事詳細文を表示する形式を用いる。

- 記事見出し文一覧画面

アプリケーション起動時の画面。記事見出し文の一覧を表示する。記事見出し文をタッチすると記事詳細画面に遷移する。また、戻るボタンを押すとアプリケーションが終了する。

- 記事詳細画面

記事見出し文をタッチしたときに表示される画面。記事見出し文に対応する記事の詳細を表示する。戻るボタンを押すと記事見出し文一覧画面に遷移する。

2.3 記事見出し文に対する興味の類推

各記事に対して、その記事の詳細を見るためにタッチしたかどうか、また、見出しに目を通した回数（以下「視認回数」という）は何回であるかという 2 点に基づいて、表 1 に示す推定ルールに基づき、興味を類推する。

ユーザが目を通したか否かは、アプリを起動してから終了するまでの間に、記事見出し文一覧画面が表示されていた時間から判定する。記事見出し文一覧の表示時間を X 、各記事を読むのに必要な時間を Y とすると、表示

表 1: 記事見出し文に対する興味の類推

興味の類推	対象とする記事見出し文
興味あり	タッチされた見出し文
興味なし	視認回数が 2 回かつタッチされなかった記事見出し文
不明	視認回数が 0 回または 1 回かつタッチされなかった記事見出し文

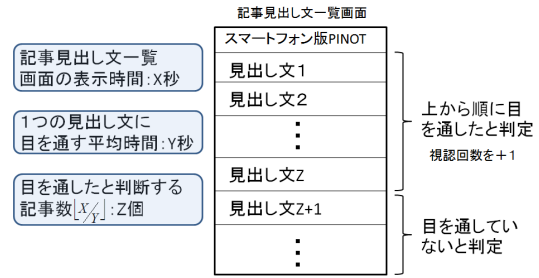


図 3: 記事見出し文に目を通したか否かの判定

中に目を通すことができた記事の数 Z は、 $Z = \lfloor X/Y \rfloor$ となる。また、ユーザは記事見出し文一覧画面において、上に表示されている記事見出し文から順に読んでいく可能性が高いと考えられる。そのため、記事見出し文一覧画面の一番上に表示されている記事見出し文から、 Z 番目の記事見出し文までを目を通した見出し文と判定し、それらの記事見出し文のそれぞれに対して、視認回数を 1 回増やす。なお、1 件の記事見出し文に目を通す平均時間は、後述の通り実験的に定めた。

1 件の記事見出し文にユーザが目を通す時間の計測

記事見出し文一覧画面における、1 件の記事見出し文にユーザが目を通す平均時間を調べるために、1 人のユーザが記事見出し文に目を通す時間を計測する実験を行った。以下に実験の条件を示す。

- 1 回の計測に用いる記事見出し文は 50 件とし、異なる 50 件の記事見出し文を対象に 3 回の評価を行う。
- 50 件の記事に対して、速読を意識して見た場合と、熟読を意識した場合の時間を計測する。
- 記事見出し文は Yahoo!ニュース [2] で配信されているニュースの見出し文を用いる

記事見出し文 50 件に目を通す平均時間は 78.5 秒であった。ゆえに、1 つの記事見出し文に目を通す時間を 1.57 秒とする。

表 2: 記事見出し文 50 件に目を通す実験の計測結果

読む速さ	回			
	1 回目	2 回目	3 回目	平均
速読	55.6s	50.9s	57.4s	54.6s
熟読	104.8s	96.7s	105.6s	102.4s
平均	80.2s	73.8s	81.5s	78.5s

2.4 ユーザプロファイルの更新

記事見出し文 W に対する興味の類推結果を基に、記事見出し文 W から抽出した各単語 w_n の新たな興味の度合い $i(w_n)$ をそれぞれ計算し、ユーザプロファイルに登録されている各単語 w_n の興味の度合い $i(w_n)$ をそれぞれ書き換える。興味の度合いの計算は、各単語 $w_n (n = 1, 2, \dots, N)$ に対して、式 (1) を適用して行う。

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

J は見出し文に対する興味の類推において、「興味有り」と判定されたなら 1、「興味無し」と判定されたなら 0 とする。単語の興味の度合いをどの程度変化させるかは、 α によって調節する。 α が小さければ、新しい興味の度合いの割合が大きくなり、逆に α が大きければ、古い興味の割合が大きくなる。

3 評価実験

興味の類推の精度を調べるため、記事見出し文に対する類推した興味とユーザの興味を取得する実験を行う。次に、興味の類推が情報フィルタリングにどのような影響を与えるか検証するため、興味の類推を行い作成したユーザプロファイルと興味の有無をユーザに直接問いかけて取得し作成した理想的なユーザプロファイルを用いて、情報フィルタリングを行った結果の比較を行う。

3.1 興味の類推の精度

興味の類推の精度を調べるために、3 人のユーザにスマートフォン版 PINOT を使用してもらう実証実験を行った。以下に実験の条件を示す。

- 配信するニュースは Yahoo!ニュース [2] で配信されているニュースを用いる
- 各ユーザのスマートフォンに配信された記事数が 1000 件になるまで使用してもらう
- 使用してもらった後、各ユーザに、配信された 1000 件の見出し文に対して興味の有無を入力してもらう

興味の類推の精度をまとめた表 3 から、興味ありと類推した見出し文のうち平均で約 7 割、興味なしと類推した見出し文のうち 9 割以上が正しく類推できていることがわかった。このことから、興味の類推は概ね正しく行えているといえる。

表 3: 興味の類推の精度

	ユーザ A	ユーザ B	ユーザ C
興味ありと類推した見出し文のうちユーザの興味ありの見出し文の割合	68.63%	53.92%	88.37%
興味なしと類推した見出し文のうちユーザの興味なしの見出し文の割合	94.06%	90.37%	96.31%

3.2 興味の類推と情報フィルタリングの関係

興味の類推と情報フィルタリングの関係を調べるため、3 名のユーザを対象に、Yahoo!ニュース [2] で配信されている記事情報を用いて実験を行った。本実験において、ユーザプロファイルを更新する際の α は 0.5、記事見出し文を表示する際の閾値は 0.75 とする。

1. ユーザプロファイルの作成

ユーザに、配信された記事数が 1000 件になるまで利用してもらい、アプリで興味の類推を行い、ユーザプロファイルを作成する。作成されたユーザプロファイルを「類推ユーザプロファイル」と呼ぶ。また、興味の類推が 100 % 正確な場合と比較するため、類推ユーザプロファイルの作成に使用した記事見出し文の興味をユーザに直接問いかけて取得し、その興味を反映したユーザプロファイルを作成する。直接取得した興味で作成したユーザプロファイルを「直接取得ユーザプロファイル」と呼ぶ。

2. 記事見出し文に対するユーザの興味の有無を直接取得

情報フィルタリングに使用する 640 件の記事見出し文に対する興味をユーザに直接問いかけて、興味の有無の判断をしてもらう。

3. 記事見出し文に対する情報フィルタリングの実施

1 で作成したそれぞれのユーザプロファイルと、2 で用いた 640 件の記事見出し文を用いて、情報フィルタリングを行う。

4. 評価

ユーザプロファイルを用いない場合、類推ユーザプロファイルを用いた場合、直接取得ユーザプロファイルを用いた場合を比較し、評価する。

3.3 評価尺度

実験の評価尺度として、情報 S/N 比、除去率、再現率を以下のように定義する。

- 情報 S/N 比

$$\text{情報 S/N 比 (\%)} = \frac{\text{表示された記事見出し文のうち、実際に興味がある記事見出し文の件数}}{\text{表示されたすべての記事見出し文の件数}} \times 100 \quad (2)$$

情報フィルタリングの際に、ユーザにとって興味があると判定し表示した全ての記事見出し文の中で、ユーザが実際に興味があった記事見出し文の割合。

- 除去率

$$\text{除去率 (\%)} = \frac{\text{実際に興味のない記事見出し文のうち、表示されなかった記事見出し文の件数}}{\text{実際に興味のない記事見出し文の件数}} \times 100 \quad (3)$$

情報フィルタリングの際に、ユーザにとって興味のない記事見出し文の中で、ユーザにとって興味がないと判定され表示されなかった割合。

- 再現率

$$\text{再現率 (\%)} = \frac{\text{実際に興味のある記事見出し文のうち、表示された記事見出し文の件数}}{\text{実際に興味のある記事見出し文の件数}} \times 100 \quad (4)$$

情報フィルタリングの際に、ユーザにとって興味のある記事見出し文の中で、ユーザにとって興味があると判定され表示された割合。

3.4 実験結果と評価

各ユーザの評価基準値をまとめた表 5 から、類推ユーザプロフィールを用いた場合、ユーザプロフィールを用いない場合と比較して、ユーザの興味のない記事見出し文を減らすことができるが、ユーザの興味のある記事見出し文も減らしてしまうことを確認した。また、類推ユーザプロフィールを用いた場合と直接取得ユーザプロフィールを用いた場合を比較すると、ユーザ A, B において、いずれの評価基準値も似た結果が得られた。従って、ユーザプロフィールの更新回数が 400~500 回の時点では、表 3 程度の興味の類推の精度であれば、100% 正確に類推ができていない場合と比較しても、情報フィルタリングに及ぼす影響に殆ど差はないと考えられる。

表 4: 類推ユーザプロフィールの更新回数

	ユーザ A	ユーザ B	ユーザ C
類推ユーザプロフィールの更新回数	405 回	487 回	536 回

表 5: ユーザ毎の評価基準

(a) ユーザ A

	情報 S/N 比	除去率	再現率
ユーザプロフィールなし	13.75%	0%	100%
類推ユーザプロフィール	16.38%	38.95%	75.00%
直接取得ユーザプロフィール	16.46%	37.50%	77.27%

(b) ユーザ B

	情報 S/N 比	除去率	再現率
ユーザプロフィールなし	19.06%	0%	100%
類推ユーザプロフィール	23.04%	35.52%	81.97%
直接取得ユーザプロフィール	22.84%	38.03%	77.87%

(c) ユーザ C

	情報 S/N 比	除去率	再現率
ユーザプロフィールなし	13.91%	0%	100%
類推ユーザプロフィール	15.89%	51.00%	57.30%
直接取得ユーザプロフィール	14.80%	35.21%	69.66%

4 結論

本研究では、個人向けにニュースを自動で選択し表示するニュースアプリの開発を行った。実験結果から、興味の類推の精度が 7~9 割で、この程度の精度であれば、理想的な条件である類推の精度が 100% の場合と比較しても、情報フィルタリングが達成する情報 S/N 比、除去率、再現率への違いは殆ど無い。しかし、ユーザプロフィールの更新回数が 500 回程度では、情報フィルタリングにおける学習が完了していないと思われる。今後、ユーザプロフィールの更新回数を増やしたときの評価を行う必要がある。

参考文献

- [1] 個人向け情報配信システム「Pinot」, <http://koblab.cs.ehime-u.ac.jp/misc/pinot/intro.html>, (参照 2015-6-23)
- [2] Yahoo! ニュース, <http://news.yahoo.co.jp/>, (参照 2015-6-23)