Vol.2016-DPS-166 No.36
Vol.2016-CSEC-72 No.36
2016/3/4

# フリーメールの広告推薦モデルにおける プライバシー向上に関する一提案

グエンスアンギア[†1]　藤田真浩[†2]　西垣正勝[†2]

**概要**：本稿では，現状のフリーメール広告推薦モデルと比較して，(i) ユーザのプライバシー向上を実現する，(ii) 現状の広告の推薦精度をできるだけ維持する，という二つの要件を両立するフリーメール広告推薦モデルの提案を行う．現在のフリーメールサービスでは，サービスプロバイダ（メールサーバ）がメール本文中に含まれる単語を分析して広告の推薦を行っている場合が多いと考えられる．そこで提案モデルでは，送信者がメール本文から名詞のみを抽出して広告推薦用のワードリストを作成し，暗号化したメール本文とワードリストのセットをフリーメールサービスのメールサーバへ送信する．メールサーバは受け取ったワードリストに現状の広告推薦アルゴリズムを適用して，受信者へ広告の推薦を行う．暗号化したメール本文は，暗号化されたままの形でメールサーバから受信者に届けられる（復号鍵は予め送信者から受信者のみに別途送られている）．本方式であれば，メールの内容すべてをサービスプロバイダへ提供していないため，ユーザのプライバシー向上が実現される．加えて，サービスプロバイダは現状の広告推薦アルゴリズムをそのまま適用できるため，広告推薦精度も維持されることが期待される．

**キーワード**：プライバシー，メール広告，フリーメール，広告推薦モデル

# An idea on advertising selection model for privacy preservation of free emails

NGUYEN XUAN NGHIA[†1]　MASAHIRO FUJITA[†2]
MASAKATSU NISHIGAKI[†2]

***Abstract***: In this paper, we propose an idea on advertising selection model to improve user privacy while maintaining the recommendation accuracy of the current free email advertising selection as much as possible. It is considered that most of the current free email services select recommended targeting ads by analyzing the words contained in the email text. In our proposed model, an email sender create a “word list” to be used for advertising selection by extracting only nouns from the email text and send it with an “encrypted email text” to a free email service. The free email service provider (the email server) can derive recommended ads to an email recipient by applying the word list to the current advertising selection algorithm. The encrypted email text is delivered to the recipient via the email server (the decryption key has been shared in advance between the email sender and recipient). This method provide only the nouns of emails to the free email service provider, and thus we consider that the user privacy is improved. This method can use the existing advertising selection algorithm, and therefore it is expected that the advertising recommendation accuracy is kept.

***Keywords***: Privacy, Email advertising, Free email, Advertising selection model

## 1. はじめに

近年，多くのフリーメールサービス（以降，単に「サービス」と呼ぶ）では，Web メールをユーザへ提供している．Web メールでは，ユーザが受信メールを閲覧するページの一部に広告を提示する場合が多い．この広告収入はサービスプロバイダにとって，メールサービスを無料で提供するための重要な資金源となっている[1]．広告提示の際に，単にランダムな広告を提示するのではなく，ユーザ（閲覧者）にそのユーザにとってより関連性が高い広告を提示する技術として「行動ターゲティング広告」が知られている．行動ターゲティング広告は，ユーザの Web ページ閲覧履歴，検索履歴，メール内容等をサーバが自動的に分析し，ユーザのプロファイル（年代，趣味，嗜好など）を把握することで，ユーザにより関連性の高い広告を提供する技術である． Gmail や Yahoo Mail!は，Web メール上で広告を提示する際にこの技術を活用していることをポリシーに明示している[2][3]．しかし，行動ターゲティング広告で利用される情報のうち，メール内容には機密情報やユーザの個人情報が含まれる場合も少なくない．したがって，行動ターゲティング広告のためにメール内容を分析されることに嫌悪感を覚えるユーザも少なくない．また，メール内容の分析は，通信秘密の観点からプライバシーの侵害になり得るという法律上の指摘もなされている[4][5]．

サービスプロバイダは，行動ターゲティングに嫌悪感を覚えるユーザに対して，メール内容の解析をオプトアウト

†1 静岡大学大学院情報学研究科
Graduate School of Informatics, Shizuoka University
†2 静岡大学創造科学技術大学院
Graduate School of Science and Technology, Shizuoka University

Vol.2016-DPS-166 No.36
Vol.2016-CSEC-72 No.36
2016/3/4

するという選択肢を与える場合が多い[3][6][7]．この対策は，ユーザのプライバシー向上の観点からは有効な対策であるが広告の推薦精度が下がってしまうため（ユーザへ関連性が高い広告が提示できなくなってしまうため），サービスプロバイダの収入減少へとつながってしまう．

そこで本研究では，現状の Web メールにおける行動ターゲティング広告を用いた広告推薦モデルと比較して，(i) 広告推薦精度をできるだけ維持しつつ，(ii) ユーザのプライバシーを向上することのできる広告推薦モデルについて検討する．本稿では，その第一歩として，送信者がメール本文から名詞のみを抽出して広告推薦用のワードリストを作成し，暗号化したメール本文とワードリストのセットをフリーメールのサービスプロバイダ（メールサーバ）へ送信する方法を提案する．メールサーバは受け取ったワードリストに現状の広告推薦アルゴリズムを適用して，受信者へ広告の推薦を行う．暗号化したメール本文は，暗号化されたままの形でメールサーバから受信者に届けられる（復号鍵は予め送信者から受信者のみに別途送られている）．

本稿の構成は次のとおりである．2 章で現状の行動ターゲティング広告の広告推薦モデルを分析する．3 章で提案方式を説明する．4 章で提案方式の広告推薦精度を，5 章でプライバシー向上に関する評価を行う．6 章で考察を行った後，7 章で関連研究を述べる．最後に，8 章で今後の課題をまとめる．

## 2. 行動ターゲティング広告の広告推薦モデル

筆者らが調べた限り，広告推薦の仕組みを詳細に公開しているフリーメールサービスはなかった[a]．そこで筆者らは，行動ターゲティング広告の仕組みについて触れられている文献[8][9][10][11]を精査し，一般的な行動ターゲティング広告の広告推薦モデルを推測することとした．

推測された「行動ターゲティング広告の広告推薦モデル」を図 1 に示す．ユーザへ推薦される広告は，各ユーザのプロファイル（年代，性別，趣味，嗜好等）から決定される．サービスプロバイダは，各ユーザのプロファイルをそのユーザの行動履歴や検索履歴等からあらかじめ作成して，Cookie ID と紐づけてプロファイル DB へ保存しておく．受信者がメールを閲覧（あるいは，受信）した際には，このプロファイル DB に保存された「受信者のプロファイル」と「そのメールの内容（タイトル・本文）」を広告推薦アルゴリズムへ入力し，受信者へ推薦すべき広告を決定する．このとき利用している「メール内容」にも受信者（ユーザ）に関する情報が含まれるため，そのメールの内容を利用してプロファイルを更新する場合もあると考えられる．

上記過程で決定された広告を受信者はメール本文と共

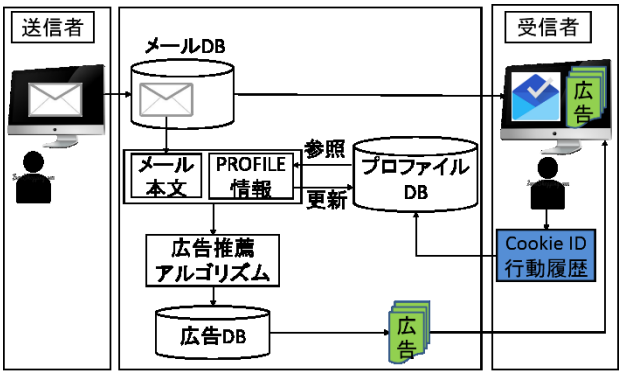


**図 1. 行動ターゲティング広告の現状モデル（推測図）**

に閲覧することとなる．以下，本稿ではこのモデルが実際に利用されているという前提のもとで，議論を進める．

## 3. プライバシー向上と広告推薦精度維持を実現する広告推薦モデル

### 3.1 コンセプト

文章の意味の多くを名詞が担うことは想像に難くない[14]．すなわち，サービスプロバイダが用いる広告推薦アルゴリズムでは，メール本文中に含まれる名詞を抽出した後，それをベースに広告の推薦を行っている場合が多いと考えられる[a]．そこで，送信者がメール本文から名詞のみを抽出して広告推薦用の「ワードリスト」を作成し，暗号化したメール本文とワードリストのセットをフリーメールのサービスプロバイダ（メールサーバ）へ送信する．メールサーバは受け取ったワードリストに現状の広告推薦アルゴリズムを適用して，受信者へ広告の推薦を行う．暗号化したメール本文は，暗号化されたままの形でメールサーバから受信者に届けられる（復号鍵は予め送信者から受信者のみに別途送られている）．

本方式であれば，メールの内容すべてをサービスプロバイダ（メールサーバ）へ提供していないため，ユーザのプライバシー向上が実現される．加えて，サービスプロバイダは現状の広告推薦アルゴリズムをそのまま適用できる可能性が高いため，広告推薦精度も維持されることが期待される．

### 3.2 提案モデル

3.1 節に示したコンセプトを，現状の広告推薦モデル（図 1）に適用することによって，図 2 に示すようなモデルが得られる．枠組み自体は現状のモデルから変更されていないことに注意されたい．メールの送信および受信の具体的な流れを以下に示す．

【送信】

① 送信者は，通常どおり，メールを作成した後，メール

a 2 章に示したとおり，筆者らが調べた限り，Web メールにおける広告推薦の仕組みを詳細に公開しているフリーメールサービスはなかった．ただし，Yahoo! Mail では，ポリシーに「メールのタイトルおよび本文を機械的に解析し，あらかじめ選んだ単語と同じ単語が含まれているかどうかを機械的に解析する」という記載がなされていた[2]．広告の目的が何らかの商品を薦めることに鑑みれば，ここでいう「単語」は名詞に中心に構成されていることが推測される．

Vol.2016-DPS-166 No.36
Vol.2016-CSEC-72 No.36
2016/3/4

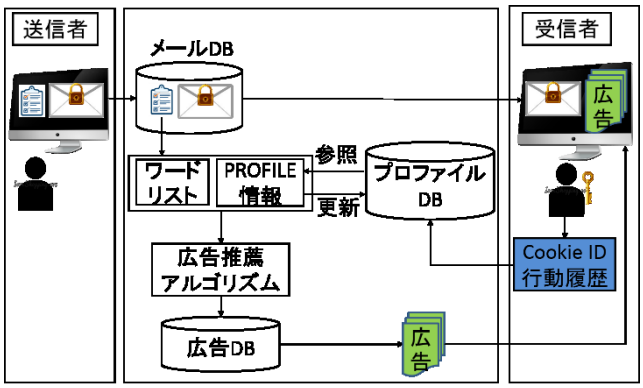


**図 2 提案モデル**

の本文を抽出する

② 送信者は，①のメール本文から広告用のワードリストを作成し，メール本文をワードリストに書き換える．

③ 送信者は，①のメール本文を暗号化し，添付ファイルとしてメールに添付する．

④ 送信者は，①～③の手順によって作成した「本文がワードリスト，添付ファイルが“暗号化された元のメール本文”」であるメールをサーバへ送信する．

【受信】

① 受信者は，自分宛に届いたメールを閲覧する．メールは，上記の送信手順①～④に従って送られている．

② メールサーバは，メール本文（ワードリスト）と受信者のプロファイルから受信者へ推薦する広告を選択する．（必要に応じて）同時に，受信者のプロファイルを更新する[b]．

③ メールサーバは，メール本文（ワードリスト）と添付ファイル（暗号化された元のメール本文）を，②で選択した広告とともにユーザへ提示する．

④ ユーザは，受け取ったメールの添付ファイル（暗号化された元のメール本文）を復号し，元のメール本文を入手する．（メール本文（ワードリスト）は，破棄してよい．）

### 3.3 ワードリストの作成方法

ワードリストの作成方法は種々の形式が考えられる．今回は，以下の 4 つの基本的な形式を用いることとした（図 3，図 4 も参照されたい）．メール本文の情報量と比較した際，形式 1，形式 2，形式 3，形式 4 の順で情報量が小さくなっていることに注意されたい．

- 形式 1（メールから名詞抽出）：
  メール本文に出現する名詞をすべて抽出し，出現した順番で列挙する
- 形式 2（メールから名詞抽出＋ソート）：
  メール本文に出現する名詞をすべて抽出し，順番を 50 音順昇順でソートした後，列挙する
- 形式 3（メールから名詞抽出＋重複削除）：
  メール本文に出現する名詞をすべて抽出し，出現した順番で列挙する．ただし，同じ名詞が複数回出現する場合，最初の 1 回のみ列挙する．
- 形式 4（メールから名詞抽出＋重複削除＋ソート）：
  メール本文に出現する名詞をすべて抽出し，順番を 50 音順昇順でソートした後，列挙する．ただし，同じ名詞が複数回出現する場合，最初の 1 回のみ列挙する．

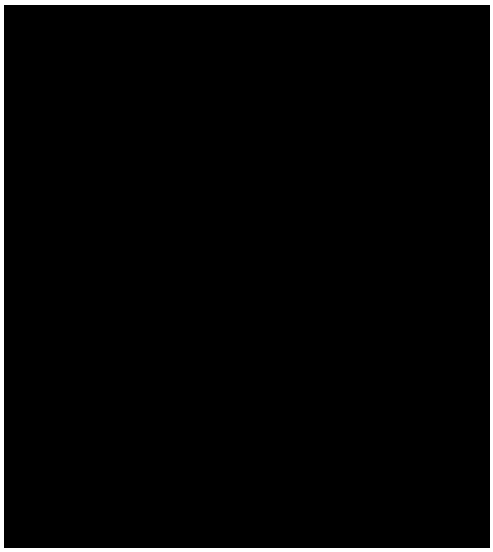

**図 3 ワードリストの作成方法**

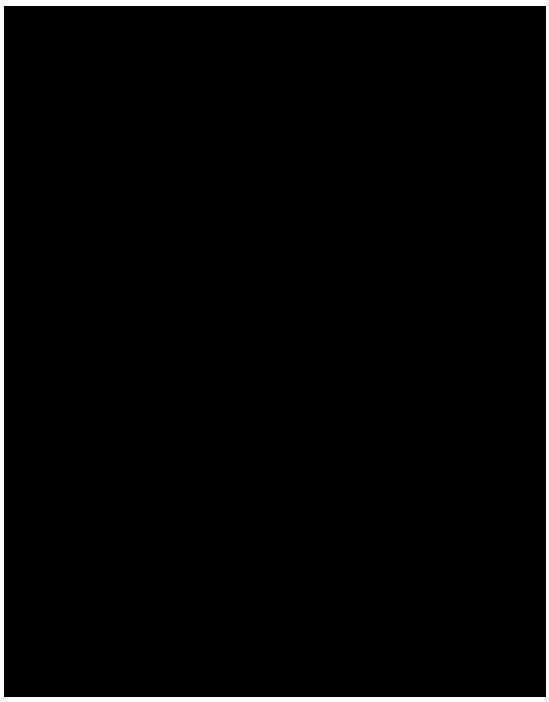

**図 4 ワードリストの例**

b これらの作業は，あらかじめ（たとえば，サーバがメールを受け取った時）に行っておいてもよい．

Vol.2016-DPS-166 No.36
Vol.2016-CSEC-72 No.36
2016/3/4

広告の推薦精度とプライバシー向上の程度はそれぞれのワードリスト形式によって異なるものと推測される．これらの評価については 4 章，5 章にて行う．

## 4. 提案モデルの広告推薦精度に対する評価

### 4.1 目的

提案モデルにおける（ワードリストから広告を推薦する際の）広告推薦精度と現在における（メールの内容すべてから広告を推薦する際の）広告推薦精度と比較する．また，用いるワードリストの種類によって，広告推薦精度がどのように変化するかを調査する．

### 4.2 方法

3.1 節に示したとおり，筆者らが調べた限り Web メールにおける広告推薦アルゴリズムを公開しているサービスはなかった．そこで本稿では，2.1 節で示した広告推薦モデルが利用されているものと仮定する．その仮定のもと，Yahoo Mail へ提案モデルを適用した際（メール本文にワードリストを記載したメールを閲覧した際）に，表示される広告を確認するという手段を用いて広告推薦精度を評価することとした．

#### 4.2.1 比較方法

提案モデルの広告推薦精度は「メール本文をそのまま送ったメール」と 3.2 節に示したモデルにおける「メール本文をワードリスト化したメール」を受け取った際に表示される広告をそれぞれ収集し，それらを比較することによって評価する．具体的な手順は下記のとおりである．

【広告収集手順】

① VM VirtualBox [12]上で新規の仮想マシンを作成する．
② ①で作成した仮想マシン上のブラウザ（Internet Explorer）で，Yahoo Mail の新規アカウントを作成する．アカウント登録時に入力が要求される個人情報は筆者に関わる情報を用いた．
③ ②のアカウントからログアウトとする．
④ ②で作成した Yahoo Mail のアカウントに対し，N 通の受信メール $M_1,M_2,\ldots,M_N$ をセットする（N 通のメールのデータを IMAP でメールスプールにコピーする）．
⑤ 以下の操作を L 回繰り返す．
 (ア) ②のアカウントにログインする．
 (イ) $M_1,M_2,M_3,\ldots,M_N$ の順でメールを順番に閲覧する．
 (ウ) Yahoo Mail には，1 つの閲覧画面につき 12 個の広告が表示される．k（1≦k≦L）回目の操作における $M_i$（1≦i≦N）を閲覧した際に表示された p（1≦p≦12）番目の広告を，$A_{k,i,p}$ とする．この結果，k 回目（1≦k≦L）の操作における $M_i$ に対しては，$Alist_{k,i}=(A_{k,i,1}, A_{k,i,2}, \ldots, A_{k,i,12})$ という「広告リスト」が得られる．
 (エ) アカウントからログアウトする．

上記手順においては，メール情報以外のプロファイルが行動ターゲティング広告に影響を与えないよう，①仮想マシンの利用や②登録情報を同一とする作業を行っていることに注意されたい．また，⑤(ウ)に示したとおり，Yahoo Mail では画面上に表示される広告は 12 個に限られている．推薦される広告が 13 個以上見つかった場合は，その中からランダムに 12 個が選ばれて表示されている可能性があるため，⑤で各メールに対し L 回の操作を繰り替えしていることに注意されたい．

【広告リスト間の類似度計算手順】

広告リスト $Alist = (A_1, \ldots, A_i, \ldots, A_{12})$が $A'list = (A'_1, \ldots, A'_j, \ldots, A'_{12})$に変化したとする．このとき，Alist と A'list の類似度 $Sim_{list}(Alist,A'list)$を式(1)のとおり定義する．

$$Sim_{list}(Alist, A'list) = \frac{1}{12} \times \sum_{j=1}^{12} \begin{cases} 0 & \text{if } A_i \neq A'_j, \forall i \\ 1 & \text{else} \end{cases} \quad (1)$$

ここで，式(1)で 1/12 を乗じているのは，一致している度数を正規化して類似度として扱うためである．

① 広告収集手順の実施によって，$M_1,M_2,M_3,\ldots,M_N$ に対する広告リスト $Alist_{k1,i}$（1≦k1≦L）を得る
② 「$M_1,M_2,M_3,\ldots,M_N$ から抽出したワードリスト」を本文としたメールを $M'_1,M'_2,M'_3,\ldots,M'_N$ とする．広告収集手順の実施によって，（$M_1,M_2,M_3,\ldots,M_N$ そのものではなく）$M'_1,M'_2,M'_3,\ldots,M'_N$ に対する広告リスト $A'list_{k2,i}$（1≦k2≦L）を得る．
③ $M_1,M_2,M_3,\ldots,M_N$ に対するすべての広告リストと $M'_1,M'_2,M'_3,\ldots,M'_N$ に対するすべての広告リストの類似度を式(2)のとおり定義する．この類似度によって，広告推薦精度は評価される．

$$\frac{\sum_{k2=1}^{L}\sum_{k1=1}^{L}\sum_{i=1}^{N} Sim_{list}(Alist_{k1,i}, A'list_{k2,i})}{L \times L \times N} \quad (2)$$

#### 4.2.2 ワードリストの作成方法

本実験で利用するワードリストは 3.3 節に示した 4 つの形式のワードリストである．4 つのワードリストは，以下の手順で動作するプログラムを用いて作成した．プログラムは Python を用いて作成されている．

① メール本文を入力する．
② ①で入力したメール本文から MECAB [13]を使って名詞をすべて抽出して配列に保存する．MECAB で使用する辞書は IPA 辞書である．
③ ②で抽出した名詞配列に対して，
 (ア) 形式 1（メールから名詞抽出）の場合：
 先頭から順にすべて出力をする．
 (イ) 形式 2（メールから名詞抽出＋ソート）の場合：
 sort 関数を用いて昇順に配列をソートする．その後，配列を先頭から順にすべて出力する．
 (ウ) 形式 3（メールから名詞抽出＋重複削除）：
 配列を先頭から順に走査する．ある要素を参照した際，その要素の値がそれ以前の要素で出現

Vol.2016-DPS-166 No.36
Vol.2016-CSEC-72 No.36
2016/3/4

している場合，ある要素を削除する．その後，配列を先頭から順にすべて出力する．

(エ) 形式 4（メールから名詞抽出＋重複削除＋ソート）：
配列を先頭から順に走査する．ある要素を参照した際，その要素の値がそれ以前の要素で出現している場合，ある要素を削除する．その後の手順は形式 2 と同様である．

#### 4.2.3 予備実験

実験前に予備実験を行ったところ，4.2 節に示した手順で類似度を求めたとき，$M_1=M'_1, M_2=M'_2, \ldots, M_N=M'_N$ であったとしても，式(2)の類似度は 1.0 にならないことが判明した．すなわち，全く同じメールを受け取っている異なるアカウントを想定した場合，それらのアカウントで同じ内容のメールを，同じ手順で閲覧したとしても，それぞれで得られる広告リストは一部が異なる（異なる広告が推薦されている）ことが明らかになった．

今回の実験においては，4.2 項で示したとおり，メール以外の情報が極力，行動ターゲティング広告に影響を与えないように，実験条件を統制している．したがって，予備実験の結果（メール情報が同じにも関わらず，異なる広告が表示されている）に鑑みて，現状の広告の推薦モデルには何らかのランダムネスが含まれているものと推測される．そこで，実験に先立って，広告リストの中にランダムに提示される広告がどの程度含まれるかを調査した．具体的には，下記の手順で調査を行った．

① N 通のメール $M_1,M_2,\ldots,M_N$ に対して「広告収集手順」を用いて広告リストを収集する．その結果，Alist が収集される．

② ①と同じ N 通のメール $M_1,M_2,\ldots,M_N$ に対して「広告収集手順」を用いて広告リストを収集する．その結果，A'list が収集される．

③ Alist と A'list についてこれらの類似度を式(2)によって計算する．

③によって得られた類似度は，（あくまで目安であるが）現在の広告推薦モデルに含まれているランダムネスの程度の基準となる．

上記手順で実験を行ったころ，得られた類似度は 0.48 であった．すなわち，広告リストで提示されるうちの広告のうち約半分は，ランダムに表示されていること（約半分の広告のみがメールの内容に依存して推薦されていること）を確認した．

### 4.3 諸元

実験で使用したメールは筆者が日常的に使用しているメールアカウントで 2015 年 9 月 1 日から 2015 年 12 月 7 日の間で受信したメール計 344 通である．すなわち，4.2 節に示した手順①，手順②においては N=344 となる．また，⑤において繰り返す回数 L は 5 に設定した．

表 1　広告推薦精度

| 形式 | 類似度 |
|---|---|
| 形式 1 | 0.47 |
| 形式 2 | 0.51 |
| 形式 3 | 0.52 |
| 形式 4 | 0.55 |

### 4.4 結果

ワードリストの形式 1～4 を利用した場合について，推薦された広告が元のメールの広告と比較してどの程度類似しているかを式(2)によって計算した．その結果を表 1 に示す．

表 1 の結果より，各形式の類似度は 0.47～0.55 である．4.2.3 項で行った予備実験の結果より，約半分はランダムな広告が表示されることがわかっている．したがって，今回得られた 0.47～0.55 という類似度は，「元のメールを送受信する方法と各ワードリストを送受信する方法を比較したところ，それぞれにおいて，ほぼ同じ広告が出現している」ことを意味している．すなわち，提案方式（ワードリストを利用した場合）においても，広告の推薦精度は現状のまま維持されることが期待できる．

一方で，形式 1～4 の間で類似度の値を比較した際，それらの間に顕著な差は見られなかった．厳密な広告推薦アルゴリズムが公開されていないため理由は不明であるが，一つの可能性としては，行動ターゲティング広告にメールを利用する際には，名詞のみ（形式 1）をみていることが考えられる．この点については，今後，さらなる文献調査を行う等して，検討を進めていきたい．

## 5. 提案モデルのプライバシー向上に関する評価

### 5.1 目的

提案モデルによる広告の推薦がユーザのプライバシー向上を実現することを確認する．また，用いるワードリストの種類が，プライバシーの向上にどのような影響を与えるかを調査する．

### 5.2 方法

メールの内容からプライバシーがどの程度向上したかを客観的に評価することは困難である．したがって本稿では，現状の広告推薦モデル（メールサーバが広告推薦のためにメール本文すべてを解析する場合）に対する嫌悪感と提案モデル（メールサーバがワードリストを広告推薦のために解析する場合）に対する嫌悪感をそれぞれアンケートによって調査し，その結果を比較することで評価することとした．

アンケートの質問項目は下記のとおりである．なお，紙面の都合上，各質問項目は実際に利用した質問を要約したものである．

Vol.2016-DPS-166 No.36
Vol.2016-CSEC-72 No.36
2016/3/4

① 性別（男性・女性）
② 年代（10 代・20 代・30 代・40 代・50 代）
③ 普段，フリーメール（例えばヤフーメール，GMAIL，HOTMAIL など）を使用しているか？（はい・いいえ）
④ 行動ターゲティング広告についてどの程度知っているか？（知っている・聞いたことはあるがよくわからない・知らなかった）
⑤ フリーメールのメールサーバがメールを分析することをどう思うか？（とても嫌・嫌・やや嫌・嫌ではない）
⑥ （提案方式について，詳しく説明した後，）
(i) あなたが「ビジネスメール（機密情報を含みうるメール）」を，次の 5 つの形式で送信する場合を想定したとき，それぞれについてどう感じるか
1. 普通のメール[c]（とても嫌・嫌・やや嫌・嫌ではない）
2. メールから名詞抽出（とても嫌・嫌・やや嫌・嫌ではない）
3. メールから名詞抽出＋ソート（とても嫌・嫌・やや嫌・嫌ではない）
4. メールから名詞抽出＋重複削除（とても嫌・嫌・やや嫌・嫌ではない）
5. メールから名詞抽出＋重複削除＋ソート（とても嫌・嫌・やや嫌・嫌ではない）

(ii) あなたが「プライベートなメール（送信先以外には知られたくない情報を含みうるメール）」を，次の 5 つの形式で送信する場合を想定したとき，それぞれについてどう感じるか
（1.～5.は(i)と同様）

(iii) あなたが「プライベートなメール（第三者に内容を知られても大丈夫なメール）」を，次の 5 つの形式で送信する場合を想定したとき，それぞれについてどう感じるか
（1.～5.は(i)と同様）

### 5.3 結果

情報系企業に勤務する 15 名を対象にアンケート調査を行った．得られたアンケート結果を次に列挙する．

① 性別：「男性」14 名，「女性」1 名．
② 年代：「20 代」4 名，「30 代」5 名，「40 代」4 名，「50 代」2 名．
③ フリーメールを使用しているか：「はい」15 名，「いいえ」0 名．
④ 行動ターゲティング広告を知っているか否か：「知っている」8 名，「聞いたことがあるがわからない」2 名，「知らなかった」5 名．
⑤ メールを分析されることをどう思うか：「とても嫌」5 名，「嫌」3 名，「やや嫌」6 名，「嫌ではない」1 名．
⑥ 各ワードリストにおける提案方式に対する感じ方：(i)～(iii)の結果について表 2 にまとめた．ここで，それぞれの値は，各ユーザの回答を 4 点（とても嫌）～1 点（嫌ではない）で数値化した後，各ユーザの平均をとった値である．

表 2 アンケート結果

| 形式 | 元のメール | 形式 1 | 形式 2 | 形式 3 | 形式 4 |
|---|---|---|---|---|---|
| (i)ビジネスメール | 3.9 | 3.3 | 3.2 | 3.2 | 3.1 |
| (ii)プライベートメール（秘密） | 3.7 | 3.3 | 3.1 | 3.0 | 2.9 |
| (iii) プライベートメール（公開可） | 2.5 | 1.9 | 1.8 | 1.8 | 1.7 |
| 平均 | 3.4 | 2.8 | 2.7 | 2.7 | 2.6 |

### 5.4 考察

表 2 より，元のメール本文をサービスプロバイダ（メールサーバ）に読まれる場合，多くのユーザが「とても嫌」あるいは「嫌」と答えており，その平均は 3.4 となっている．特に，(i)ビジネスメールや(ii)プライベートメール（秘密）においては，その傾向は顕著である．以上の結果は，現状のターゲティング広告に対して，メールサーバがユーザのメール本文すべてを分析していることに一定の嫌悪感を覚えていると考えられる．

表 2 のとおり，元のメールでは 3.4 点だった嫌悪感の平均値が，形式 1～形式 4 のワードリストでは 2.6～2.8 である．この結果は提案方式がユーザのメール分析に対する嫌悪感を一定の割合で下げることに成功しているといえるだろう．すなわち提案方式は，現状の広告推薦モデルと比較して，ユーザのプライバシー向上に一定の寄与があることが確認された．

元のメール本文の情報量に対し，ワードリストは形式 1，形式 2，形式 3，形式 4 は，この順で文章の情報量が少なくなっている．表 2 においても，形式 1，形式 2，形式 3，形式 4 の順で嫌悪感は小さくなっている．すなわち，メール本文の情報量を少なくしていく操作は，ユーザの嫌悪感を下げることに一定の効果があるものと期待される．しかし，現段階では被験者数がまだ少なく，さらに，それぞれの形式間の差は非常に小さいため一般性がいえるものではない．今後，被験者数を増加して，さらなる調査をする必要があるだろう．

c 元のメール本文（メール本文そのまま）のことである．

## 6. 議論

### 6.1 メール本文の暗号化・復号

提案方式は受信者・送信者間で事前に鍵共有が必要である．ユーザ間で鍵を共有する仕組みは，公開鍵基盤（Public Key Infrastructure）や PGP（Pretty Good Privacy）等の既存の技術を利用することが可能である．暗号化・複合の処理については，Web ブラウザのプラグインとしてユーザに提供する等によって，ユーザ自身が提案方式の枠組みを意識せずに（現在の Web メールと全く同じ感覚で）利用することも可能であろう．

### 6.2 ワードリストの形式

今回は 3.3 節に示した 4 つの形式のワードリストを用いて提案方式を評価した．しかし，5 章で実験を受けた被験者の一人から，実験終了後に，以下のようなワードリスト構築方法がコメントとして寄せられた．

*例えば，フリーメールサービス主が分析の為に必要となる最小限の情報を送る，という方法があると良いと思いました．*
*例)メールは暗号化して受信者のみに送付し，メール内の頻出名詞上位 5 件を“出現回数を伏せて”，ソートした状態で SP に提供する．*

被験者からのコメントで得られたように，今回利用した形式をより発展させた形式も考えられる．今後，他のワードリスト形式についても調査していく必要がある．

### 6.3 送信者の不正

提案方式に対して，送信者は次の手順で不正が可能である．

① 送信者（不正者）は，通常どおりメールを作成した後，メールの本文を抽出する
② 送信者は，①のメールとは無関係な文章（あるいは，ワードリスト）をメール本文とする．
③ 送信者は，①で抽出した本文を暗号化した後，添付ファイルとしてメールに添付する．
④ 送信者は，②③の手順によって作成した「本文が“無関係な文章”，添付ファイルが“暗号化された元のメール本文”」であるメールをサーバへ送信する．

すなわち，サービスプロバイダ（メールサーバ）へ提供するワードリストを全く関係ない文章へと変更する不正である．その結果，サービスプロバイダが受信者に対して適切な広告を推薦できなくなったり，サービスプロバイダが受信者に対して推薦する広告を送信者が意図的に操作できたりする可能性がある．更には，ワードリストの中には適正な単語のみを記しておくことによって正当なメールを装い，暗号化された添付ファイルの中に不正なメールを仕込んで受信者に送りつけるという攻撃も考えられる．提案方式においては，元のメールの本文自体は暗号化されて送受信されるため，サービスプロバイダ（メールサーバ）側がメールに対するウィルス検査やスパム検査を実施することができない．これらの不正に対する根本的な対策は今後の課題である．簡易的な対策として，6.1 節に示した「プラグインをユーザへ提供する」環境下においては，受信者が自身の受信したメール一通々々に対して「ワードリストと添付ファイル（元のメール本文）との一貫性」を確認し，不正であった場合にはサーバへ報告するといったような対策をとることで，一定の効果が期待される．

## 7. 関連研究

筆者が調べた限り，行動ターゲティング広告のプライバシー保護を，既存のフリーメールサービスのビジネスモデルと技術的プラットフォームを維持した形で実現した研究は過去に存在していない．そこで本章では，行動ターゲティング広告におけるプライバシー保護に対する提言，および，メール分析技術に関する既存研究について一例ずつ採り上げて概説する．

### 7.1 行動ターゲティング広告におけるプライバシー保護

文献[8]では，「利用者の閲覧履歴，検索履歴等のような行動に関する様々な情報の収集および分析を行って，ユーザの興味・関心を推定し，ユーザと関連の高い広告を推薦する」という行動ターゲティング広告の特徴から，ユーザのプライシー侵害に対して警鐘を鳴らしている．そして，個人識別できないような情報の収集・提示に向けての要件を提示している．

### 7.2 メール文書からの知識抽出

文献[14]では，データベースの中から価値ある情報を発見するデータマイニング技術である KDD（Knowledge Discovery in Databases）を，電子メールアーカイブに適用した KDM（Knowledge Discovery in electronic Mail）技術が提案されている．具体的には，メールのタイトルと本文から単語頻度情報等を抽出した上で，得られた情報を関係データベースに格納し，知識獲得アルゴリズムを用いて相関ルールを導出する．文献[14]の知見は，提案方式がメールの本文から抽出した名詞を利用する理由の一つとなっている．

## 8. まとめと今後の課題

本稿では，フリーメールにおける新しい広告推薦モデルについて検討した．提案広告推薦モデルに関し，現状の広告推薦モデルと比較して，(i) 広告精度が維持されているか，および，(ii) ユーザの不満度がどう変化するかを調査した．その結果，提案モデルは (i) 現状のモデルの広告推薦精度を維持しており，(ii) ユーザのプライバシー向上に一定の寄与があることを確認した．今後の課題としては，送信者の不正に対する対策の実施，他の形式のワードリストの利

用に対する調査，提案方式のプライバシー向上に関する客観的（理論的）評価が挙げられる．

## 謝辞

本研究の 5 章評価実験においてご協力していただいた株式会社 東芝 インダストリアル ICT ソリューション社の加藤岳久様と同僚の皆様及び三菱電機株式会社 情報技術総合研究所の小林様・柴田様と同僚の皆様に感謝の意を申し上げます．

## 参考文献

1) 総務省　情報通信政策研究所：　行動ターゲティング広告の経済効果と利用者保護に関する調査研究報告書, (2010)
2) グーグルのポリシーと規約,入手先<https://www.google.co.jp/intl/ja/policies/privacy/>, 2016.02.06 参照
3) ヤフーメールのインタレストマッチ広告と設定，入手先<http://info.mail.yahoo.co.jp/im_optout/>, 2016.02.06 参照
4) 福家秀紀 :”回線開放の歴史的意義 : ビッグデータと通信の秘密の視点から”, Journal of Global Media Studies13, pp. 59-73, 2014.
5) Samir Chopra, Laurence White: “Privacy and Artificial Agents, or, Is Google Reading My Email? ”, Vol. IJCAI-2007, pp. 1245-1250 (2007).
6) Yahoo!JAPAN 行動ターゲティング広告の無効化について，入手先<http://btoptout.yahoo.co.jp/optout/preferences.html>, 2016.02.06 参照
7) グーグル広告の管理，入手先<http://www.google.com/settings/ads/anonymous>, 2016.02.06 参照
8) 西村俊介，田中英彦 : “行動ターゲティング広告におけるプライバシー保護の実現方式についての検討”，情報科学技術フォーラム講演修論集，Vol. 10, No. 4, pp. 527-528 (2011).
9) Jeffrey A.Dean, Georges R.Harik, Paul Buchheit. Serving advertisements using information associated with e-mail. U. S. Patent 20040059712A1. 2004-03-25.
10) Jamey Graham, David G. Stork. Content based web advertising. U. S. Patent 006804659B1. 2004-12-12.
11) Darrell Anderson, Paul Buchheit, Alexander Paul Carobus, Yingwei Cui, Jeffrey A. Dean, Georges R. Harik. Deepak Jindal, Narayanan Shivakumar. Serving advertisements based on content. U. S. Patent 007136875B2. 2006-11-14.
12) Oracle 社　VM VirtualBox , 入手先< https://www.virtualbox.org>, 2016.02.06 参照
13) 工藤拓, 山本薫, 松本祐次 : “Conditional Random Fields を用いた日本語形態素解析”，情報処理学会研究報告自然言語処理（NL）, Vol. 2004, No. 47, pp. 89-96 (2004).
14) 上田 宏高，柳沢豊，塚本昌彦，西尾章治郎 : “電子メールの傾向分析への知識獲得手法の適用”，情報処理学会論文誌，Vol.41,No.12, pp.3285-3294，(2000)

# 正誤表

下記の通り、図の誤りがありますので、訂正いたします。

位置：ページ 3 の図 3、図 4

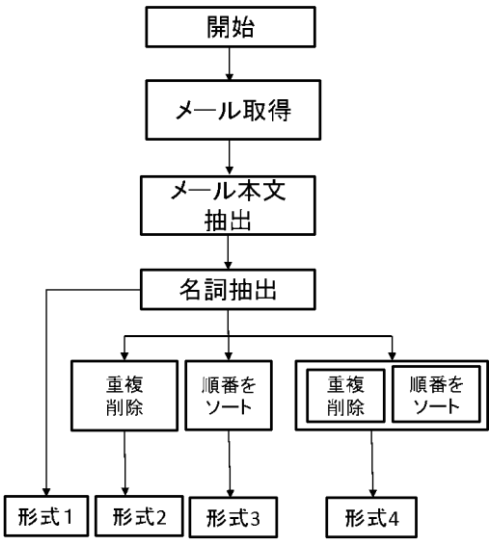


図 3 ワードリストの作成方法

**元のメール:**

こんにちは
今日は良い天気ですね。
今日は一緒にショッピングに行きませんか？

**形式1:メールから名詞抽出**

今日 天気 今日 一緒 ショッピング

**形式2:メールから名詞抽出＋ソート**

ショッピング 一緒 今日 今日 天気

**形式3:メールから名詞抽出＋重複削除**

今日 天気 一緒 ショッピング

**形式4 :メールから名詞抽出＋重複削除＋ソート**

ショッピング 一緒 今日 天気

図 4　ワードリストの例