情報処理学会論文誌 Vol.57 No.12 2731–2742 (Dec. 2016)

企業グループに送られた標的型攻撃メールのソーシャルエンジニアリング視点からの分析

渡部 正文[1,2,a)] 島 成佳[1,2,b)] 今橋 泰則[3,c)] 吉岡 克成[4,2,d)] 高木 大資[5,e)]

受付日 2016年3月10日, 採録日 2016年9月6日

概要：標的型攻撃メールには，受信者を巧みに操るソーシャルエンジニアリングが悪用されることが多く，受信者が騙しの細工に引っ掛かり端末が感染することが，情報流出などの被害に至る最初の原因である．組織の情報セキュリティ部門では，標的型攻撃メールを発見し感染を防ぐためにメールを監視しているが，騙しの細工に着目した監視は人の経験に依存しており，監視者への負担が重いことが課題となっている．筆者らは，この課題を解決するため，件名の単語を対象に，標的型攻撃メールの弁別に有効な単語や指標を明らかにする分析手法を創出する．実在する企業の協力によりメールデータセットを分析し，分析手法が標的型攻撃メールの弁別に有効であることを示す．

キーワード：標的型攻撃メール，ソーシャルエンジニアリング，統計分析，サイバーセキュリティ

Analysis of Targeted Attack Mail Sent to an Enterprise Group from Social Engineering Point of View

Masafumi Watanabe[1,2,a)] Shigeyoshi Shima[1,2,b)] Yasunori Imahashi[3,c)]
Katsunari Yoshioka[4,2,d)] Daisuke Takagi[5,e)]

Received: March 10, 2016, Accepted: September 6, 2016

Abstract: In targeted attack mail, social engineering is often used to manipulate recipients psychologically. The first cause of information leakage is that the recipients are tricked by word and phrase and the recipients' PCs are infected. Information security divisions in organizations monitor email in order to find targeted attack mail and stop infections. However, there are some problems that the burden to the operators is heavy because the monitoring based on words is depending on operators' experience. In order to solve the problems, the authors of this paper create analysis methods to find effective words and indices to discriminate targeted attack mail based on words appeared in subjects. Analyzing dataset with a real enterprise's cooperation, this paper shows that the analysis methods are effective to discriminate targeted attack mail.

Keywords: targeted attack mail, social engineering, statistic analysis, cyber security

[1] 日本電気株式会社セキュリティ研究所
Security Research Laboratories, NEC Corporation, Kawasaki, Kanagawa 211–8666, Japan
[2] 横浜国立大学先端科学高等研究院
Institute of Advanced Science, Yokohama National University, Yokohama, Kanagawa 240–8501, Japan
[3] 株式会社 NEC 情報システムズプラットフォームサービス事業部
Platform Services Division, NEC Informatec Systems, Ltd., Kawasaki, Kanagawa 211–8666, Japan
[4] 横浜国立大学大学院環境情報研究院
Graduate School of Environment and Information Science, Yokohama National University, Yokohama, Kanagawa 240–8501, Japan
[5] 東京大学大学院医学系研究科
Graduate School of Medicine, The University of Tokyo, Bunkyo, Tokyo 113–0033, Japan
a) watanabe@cj.jp.nec.com
b) shima@ap.jp.nec.com
c) imahashi@pb.jp.nec.com
d) yoshioka@ynu.ac.jp
e) dtakagi-utokyo@umin.ac.jp

1. はじめに

企業や官公庁を含む様々な組織に，個人情報や機密情報の窃取に利用される標的型攻撃メールが届いている．日本の警察庁の発表によると，2014 年に警察が把握した標的型攻撃メールによる攻撃は 1,723 件と前年比約 3.5 倍であり [1]，増加傾向にある．そして，標的型攻撃メールに起因する情報流出事案が繰り返し発生し，企業や官公庁などの情報セキュリティ部門では深刻な問題になっている．2015 年 6 月に日本年金機構が公表した情報流出事案では，標的

型攻撃メールを使った侵入が発端となり，最終的に基礎年金番号と氏名を含む 125 万件の個人情報が外部に流出した [2].

標的型攻撃メールは，標的型攻撃 [3] の初期段階において，攻撃者が組織内部に侵入する手段として利用される．攻撃者は，まず，人が端末を操作して開くと端末がマルウェアに感染する不正な添付ファイルや，人が端末を操作してアクセスすると端末がマルウェアに感染する不正なウェブサイトを準備する．次に，これらの，不正な添付ファイルを開いたり，不正なウェブサイトへのリンクをクリックしたりするように，組織内部の人物にメールを送る．

標的型攻撃メールには，受信者に不正な添付ファイルを開かせたり，ウェブサイトへのリンクをクリックさせたりするために，相手の心を巧みに操るソーシャルエンジニアリング [4] が悪用されていることが多い．件名や本文は，信用できる送信元からの業務上の用件を装っており，添付ファイルやウェブサイトなどを確認する必要があると思わせる様々な騙しの細工が施される．騙しの細工の例として，(1) 件名や本文に具体的な刊行物名をあげ取材の申込みの旨が書かれている，(2) 本文の冒頭に受信者の所属と本名が書かれている，(3) 開くことを動機付けるような本文や添付ファイル名である，などが報告されている [5]．受信者がこの騙しの細工にうっかり引っ掛かってしまうことが，情報流出などの被害に至る最初の原因である．標的型攻撃メールは，システムではなく人の心理の脆弱性を突いているともいえる．

受信者が不正な添付ファイルを開いたり，不正なウェブサイトへのリンクをクリックしたりして，端末がマルウェアに感染してしまうと，遠隔操作ツールのダウンロードとインストール，攻撃者の用意した中間サーバへのコールバックなどが起こる．端末は，外部からの遠隔操作や，端末中のファイルなどの覗き見だけでなく，新たな不正プログラムのインストールや，この端末を踏み台にした組織内の他の端末やサーバへの感染拡大も可能になる．そして，情報流出などの深刻な被害の発生に至る．

騙しの細工に対して，企業や官公庁などの組織の中には，受信者が注意するだけではリスクが大きいと考え，情報セキュリティ部門によるメール監視を導入しているところがある．このような組織は，メール監視により，業務メールに紛れ込む標的型攻撃メールを検知し，その配送を止めたり，受信者の注意を喚起したりして，攻撃者の侵入を防ごうとしている．

標的型攻撃メールは，検知装置で漏れなく検知することが技術的に困難なため，メール監視担当者（以降，監視者とする）による目視と組み合わせた総合的なメール監視が必要となる．特に，標的型攻撃メールの件名や本文に施される騙しの細工に基づく検知技術の研究としては有効なものが見られないため，騙しの細工に着目したメール監視は，監視者の経験に基づく目視に頼っている．

筆者らは，ある企業グループ（以降，企業グループ A とする）の情報セキュリティ部門の協力を得て，監視者にヒアリングをしたところ，監視者の経験に基づく監視には限界があることが分かった．メール監視の現場では，チェック対象にあがるメールが膨大であり，さらに，騙しの細工も 2012 年頃から巧妙化し業務と見分けがつき難くなっているため，監視者の負担が重くなっていることが分かった．また，大きな組織では 1 日 100 万のオーダを超える数のメールを受信しており，これらのメールを対象に騙しの細工に着目して監視をすることは，すでに人手では限界な状況である．

本研究は，騙しの細工に着目したメール監視作業を監視者の経験を参考に機械化し，監視者の負担を軽くすることを目指す．その第 1 歩として，筆者らは，件名の騙しの細工に着目する．本論文では，監視者へのヒアリング結果を整理して，件名の騙しの細工に着目した標的型攻撃メールの弁別に向けた課題を明らかにし，この課題を解決する手法を創出する．本手法は，実際の標的型攻撃メールのデータセットを用いて評価し，メール監視作業の機械化に有効な結果が得られたため報告する．

2. 標的型攻撃メール

2.1 標的型攻撃メールの特徴

標的型攻撃メールには次の特徴がある：

特徴 1 組織によって受け取る内容の傾向が異なる

標的型攻撃メールは，組織の個人や業務に合わせて作られるため，標的型攻撃メールの内容の傾向は，受け取る組織や人物によって異なる．件名や本文に使われている用語は，受信者の業務に関係するものが使われる．差出人は，受信者の業務に関係する実在の人物・組織・機関，あるいは，昔の同僚や同期の社員など受信者と個人的なつながりのある人物の氏名に偽装されており，メールアドレスも本物やそれときわめて似たものが使われる．

特徴 2 組織の外に提供することが難しい

標的型攻撃メールは，前項のように組織の業務に合わせて作られるため，組織固有の情報や秘密情報を含む場合があり外部提供が難しい．

2.2 標的型攻撃メールの関連研究

マルウェアに関しては，大量の事例データを収集したサービスがあり[*1]，このサービスを活用した研究がさかんに行われ，研究から有効な知見が得られている．たとえば，ファイルアクセス失敗挙動に基づく，情報探索型マルウェ

[*1] ユーザから投稿されたファイルなどを複数のウィルス対策ソフトで検査して結果を表示する VirusTotal と呼ばれるサービスがあり，ユーザから投稿された大量の情報を蓄積している．
https://www.virustotal.com/

アの検知手法の提案では，上記サービスから入手した情報漏洩を行うことが予想される実マルウェア検体に対して提案手法を適用し，情報探索型マルウェアを検知できることを示した [6]．文書ファイルに実行ファイルを埋め込んだ場合に，ファイルの構造に不整合が生じる特徴を検査することで悪性文書ファイルを高速かつ高確率で検知する手法 [7] や，マルウェア感染直後のプロセス構造に着目し，おとり文書を表示するマルウェアの検知手法 [8] を提案した研究がある．

標的型攻撃メールの研究においても事例データが重視される．しかし，標的型攻撃メールの事例データは，前節の特徴 2 で述べたように，秘密情報を含むことがあるため組織外に提供され難く，入手がきわめて困難な状況である．

行政機関やその関連団体は，企業などの組織から標的型攻撃メールの事例データを集めることができた数少ない例である．これらの機関や団体は，寄せられた事例データの記述統計をもとに考察を述べた調査報告書を発行している．しかし，事例データの詳細は公表されていない [1], [9]．また，個別の事例を示した注意喚起もあるが，数が少なく詳細部分が加工されており事例データとしての利用は難しい [5]．

このような状況の中，2014 年になって標的型攻撃メールの詳細な事例データを入手して行われた研究が見られるようになった．Blond らは，ある非政府組織の職員に送られた標的型攻撃メールの事例データを入手し，標的型攻撃メールに使われるソーシャルエンジニアリングの実態を次のように分析した [10]：

- 内容は，受信者の母国語で書かれている場合が最も多く，標的型攻撃メールは受信者の母国語に合わせて作られていた．
- 受信者の業務に関連するイベントの案内文が流用されるなど，標的型攻撃メールは受信者の専門分野に合わせて作られていた．
- 差出人メールアドレスは，受信者が知る人物のメールアドレスに類似している場合が最も多く，標的型攻撃メールは受信者が直接知る人物になりすまされていた．

このほかにも，収集した標的型攻撃メールの宛先に指定されていた組織の実名や，被害者に標的型攻撃メールが送られた時期や頻度が分析された．事例データをソーシャルエンジニアリングの観点から詳しく分析した研究はほかには見当たらない．

先行研究では，非政府組織の職員に送られた通常業務メールの事例データは分析の対象ではなかった．対象組織に届いた標的型攻撃メールと通常業務メールとを比較してメール監視作業に役に立つ指標を明らかにする研究が必要とされている．

3. 課題と解決アプローチ

本章では，標的型攻撃メールを騙しの細工に着目して弁別するための課題を明らかにし，課題の解決アプローチを述べる．課題を明らかにするためには，組織の情報セキュリティ部門におけるメール監視業務の現状を把握する必要がある．

そのために，筆者らは，標的型攻撃メールの問題をかかえている，企業グループ A の情報セキュリティ部門から協力を得て，監視者にヒアリングを行った．

3.1 メール監視業務

本節では，企業グループ A におけるメール監視業務の現状を把握する．企業グループ A におけるメール監視運用モデルの概略を図 1 に示す[*2]．従業員は，教育や訓練を受けているが，標的型攻撃メールの専門家ではなく，騙しの細工に引っかかってしまう危険性がある．このため，企業グループ A では，従業員の判断のみに頼らず，監視者も標的型攻撃メールに施される騙しの細工に着目してメールを監視している．

外部から到着したメールは，まず，(1) アンチウィルスで検査され，既知のウィルスと判定されたメールは破棄される．これを通過したメールは，次に，(2) ラベル付与において従業員がフリーメールアドレスから送られたメールなどに気づきやすくするためのラベルが付与される．次に，メールは，(3) 従業員のメールボックスに配送され，同じものが (4) 検知装置群にも配送されサンドボックスでの動的解析やヒューリスティクスを用いて添付ファイルなどが検査される．

標的型攻撃メールは，検知装置で漏れなく検知することが技術的に困難なため，監視者の目視と組み合わせた総合的なメール監視が必要となる．特に，標的型攻撃メールの件名や本文に施される騙しの細工に基づく有効な検知技術

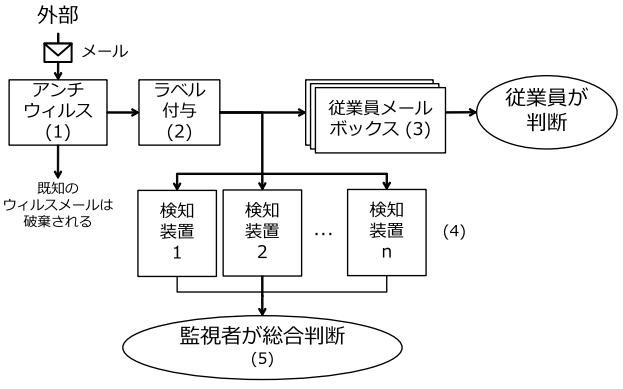

図 1 メール監視運用モデル
Fig. 1 Monitoring operation model for email.

[*2] メール監視運用の実状は組織によって異なるため，本節で述べる問題がそれぞれの組織においてもあてはまるか注意する必要がある．

の研究や実用化は見られないため，ソーシャルエンジニアリングの観点からは，監視者が，経験に基づいてメールを監視しなければならない．

企業グループ A では，1 日に 100 万のオーダを超える数のメールを受信しており，監視者がすべてのメールをソーシャルエンジニアリングの観点から目視確認することは不可能である．このため，監視者は，各検知装置が検知したメールについて上記観点を含めた目視確認をして標的型攻撃メールかどうかを（5）総合判断する．

総合判断は，まず，件名と差出人と宛先のみを確認して標的型攻撃が疑われるメールを絞り込む一次フィルタリング作業と，次に，絞り込まれたメールの本文や添付文書を詳しく調べる詳細調査作業の 2 段階に分かれており，これらのうち一次フィルタリング作業における負担が高い．

3.2 目視による一次フィルタリング作業の問題

標的型攻撃メールの見逃しを防ぐため，各検知装置は，過検知を許容した運用をしており，監視者が 1 回の一次フィルタリング作業において目視確認するメールの数は，聞いたところによると，数千を超えるオーダの数にのぼる．監視者は，限られた時間の中で，膨大な数のメールを確認しなければならない．

さらに，標的型攻撃メールの差出人や件名に施される騙しの細工は巧妙であり，通常業務メールの差出人や件名との見分けがつき難い．騙しの細工に着目した標的型攻撃メールの弁別は，現状は監視者の経験のみに基づいている．筆者らが監視者にヒアリングしたところ，次のような問題があることが分かった：

問題 1 要注意単語が含まれている件名の場合は，標的型攻撃メールが疑わしいと判断しているが，要注意単語を選択する確固たる基準がない．

問題 2 短い件名の場合は，標的型攻撃メールが疑わしいと経験的に判断しているが，件名の長さが妥当な判断基準であるかはっきりしていない．

問題 3 固有名詞が少ない件名の場合は，標的型攻撃メールが疑わしいと経験的に判断しているが，固有名詞の少なさが妥当な判断基準であるかはっきりしていない．

問題 4 珍しい件名の場合は，標的型攻撃メールが疑わしいと経験的に判断しているが，件名の珍しさが妥当な判断基準であるかははっきりしていない．

問題 5 差出人メールアドレスの「@」より後ろのドメイン名が珍しい場合は標的型攻撃メールが疑わしいと経験的に判断しているが，ドメイン名の珍しさが妥当な判断基準であるかははっきりしていない．

問題 6 宛先メールアドレスから分かる従業員の役職や所属部門を考慮して，件名や差出人メールアドレスの疑わしさを判断する必要もあり，その場合は，判断がさらに難しい．

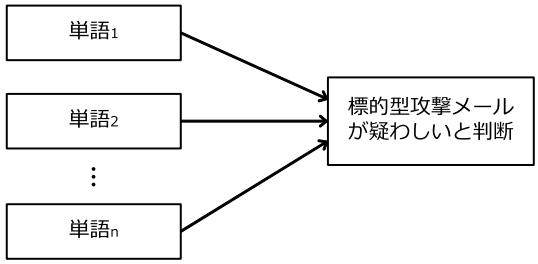

図 2 仮説モデル 1
Fig. 2 Hypothesis model 1.

3.3 課題

監視者は，限られた時間の中で，経験のみに基づいて膨大な数のメールを監視しなければならない．目視によるメール監視は限界にきている状況といえる．そこで，事例データに基づく統計学的な分析を行い，一次フィルタリング作業における監視者の経験の補完や負担の軽減に資する分析手法の開発が必要とされている．

理想的には，メールを構成する様々な要素（件名，差出人，本文，添付文書名など）を対象に騙しの細工を手掛かりにした標的型攻撃メールの弁別に取り組むことが望ましい．しかし，企業グループ A からは件名についてのみ事例データの提供の協力が得られた．利用可能なデータに制約があったが，件名は，それ自体がメールの内容を要約したり代表したりする要素であるため，一次フィルタリング作業において，標的型攻撃メールであることが疑わしいと判断する際に，監視者が最も重視している要素である．本論文では，件名のみを分析の対象とすることにした．

筆者らは，件名に関する問題 1〜4 について監視者と考察し，これらの問題を単語の視点からとらえ直し，次の 2 つの課題に整理した：

課題 1 特徴的な単語が件名に出現するか否かを基準にして標的型攻撃メールを弁別する ··· この課題を解決すると問題 1 が解決される．

課題 2 件名に含まれる単語の指標（単語数，単語の珍しさ，固有名詞率）を基準にして標的型攻撃メールを弁別する ··· この課題を解決すると問題 2〜4 が解決される．

3.4 課題解決アプローチ

筆者らは，前節の課題 1 と課題 2 を，それぞれ，以下の仮説を立てて解決することにした．

3.4.1 仮説 1

筆者らは，件名に特徴的な単語（単語$_1$, 単語$_2$, ..., 単語$_n$）が出現すると，標的型攻撃メールが疑わしいと判断できるとの仮説を置くことにした（**図 2**）．

3.4.2 仮説 2

問題 2〜4 から，監視者が標的型攻撃メールが疑わしいか否かを判断する際に参考にしている次の指標を抽出した：

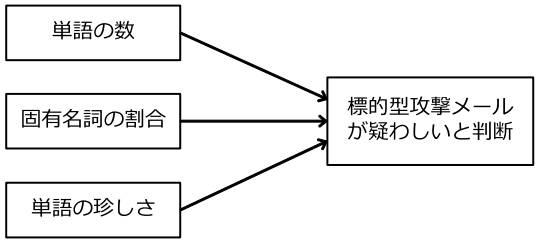

図 3 仮説モデル 2
Fig. 3 Hypothesis model 2.

(1) 単語の数
監視者は，単語数が少ない件名に違和感を感じる．
(2) 固有名詞の割合
監視者は，固有名詞が使われていない件名に違和感を感じる．
(3) 単語の珍しさ
監視者は，珍しい単語が使われている件名に違和感を感じる．

筆者らは，件名における (1) 単語の数と (2) 固有名詞の割合と (3) 単語の珍しさから，標的型攻撃メールが疑わしいと判断できるとの仮説を置くことにした（**図 3**）．

4. データセットの構築

仮説を検証するためには，組織における事例データが必要である．筆者らは，企業グループ A の関係部門の協力を得て，件名について，**表 1** のデータセットを使用した．

既存研究や調査における分析は，標的型攻撃メールのデータのみを対象とした記述統計である．しかし，本論文は，通常業務メールに紛れ込む標的型攻撃メールの弁別を目的とした分析をするため，標的型攻撃メールと通常業務メールの 2 つのデータが必要であり，この点が既存研究と異なる．

4.1 標的型攻撃メールデータセット

標的型攻撃メールデータセットは，表 1 に記載の期間（903 日間）に，外部から企業グループ A の従業員メールアドレス宛てに送られた 185 件の標的型攻撃メール（全数）を含む．標的型攻撃メールかどうかは，企業グループにおけるサイバー攻撃対策の意思決定長と複数の監視者が合議において，当該メールの詳細調査結果に加えて，受信者へのヒアリング結果，過去のログ，外部から提供された注意喚起情報などに基づいて判断した．

4.2 通常業務メールデータセット

通常業務メールデータセットは，表 1 に記載の期間（903 日間）に，外部から企業グループ A の従業員メールアドレス宛てに送られた業務メールの件名を含む．

上記期間に届いた全業務メールの件数は，数億のオーダを超えるため，本論文では，系統抽出法を用いて 10,000 件[*3]を目標に無作為抽出した．蓄積されているメールは，配送日時と時刻を指定して取得する運用のため，**図 4** に示すような各曜日・毎時における平均メール到着数モデルを作り，このモデルに基づいて抽出対象のメールの配送日時と時刻を算出し，その時刻に到着したメール（図 4 の抽出対象部分）を取得し，取得したメールが複数ある場合は，さらにその中から無作為に 1 件を選択し 11,270 件を抽出した．

表 1 メールデータセット
Table 1 Email dataset.

	標的型攻撃メールデータセット	通常業務メールデータセット
期間	2012 年 11 月 30 日～2015 年 5 月 22 日（903 日間）	
件数	185 件（全数）	11,270 件（上記期間の標的型攻撃を除く全メールから系統抽出法により無作為抽出）

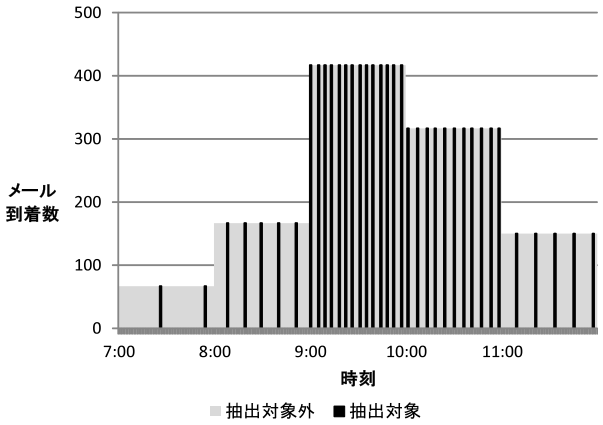

図 4 平均メール到着数モデル（例）
Fig. 4 Model of average amount of incoming email (example).

5. データセットを用いた仮説検証

本章では，3.4 節で述べた仮説を 4 章で述べたデータセットを対象に統計分析を行うことによって検証する．統計分析には，Stata SE 13.1 を使用した．以下に，分析手法の詳細と分析結果を示す．

5.1 仮説 1 の検証

疫学の分野では，発生が稀で十分な症例データが得られない疾病の要因を調査する場合に，ケースコントロール研究が広く用いられる [11], [12]．標的型攻撃メールも稀にしか受け取らない組織が多く，分析に十分な事例データを得ることは難しい．このため筆者らは，ケースコントロール研究を援用して課題 1 を解決する統計分析手法を創出することにした．

分析手法 1

本論文では，以下の統計分析手法で，標的型攻撃メール

[*3] 抽出目標件数は，企業グループ A の情報セキュリティ部門から提供が許可された件数を参考に決定した．

の弁別に有効な単語を抽出する：

(1) まず，データセットから単語の出現情報と標的型攻撃メール情報を変数値として持つ多変量データを作成する．

(2) 次に，単語の出現情報（独立変数）と標的型攻撃メール情報（従属変数）との間にロジスティック回帰モデルをあてはめ，この回帰モデルにおいて統計学的に有意な独立変数を標的型攻撃メールの弁別に有効な単語とする．

5.1.1 多変量データの作成

データセットから単語の出現情報と標的型攻撃メール情報を変数値として持つ多変量データを作成した．

件名から単語を抽出する必要がある．単語の抽出方法は，一般的に形態素解析が用いられている．今回は，テキストマイニングに広く使用されている茶筌 [13] を利用して件名を形態素に分解した．件名に多用される品詞の傾向を参考にして，得られた形態素のうち名詞，副詞および感動詞を分析の対象とした．企業グループ A に届いた標的型攻撃メールの件名は日本語で書かれているものがほとんどであるため，日本語の単語を分析対象とした．

上記のようにして，2 つのデータセットにおける合計 11,455 件の件名から，9,140 語のユニークな単語を抽出した．メールに含まれる秘密を保護するため，本論文では，具体的な単語を表記しない．これらの単語に対応する二値変数 $w_1, w_2, \ldots, w_{9140}$ を設け，件名が変数に対応する単語を含んでいる場合は，その値を 1 と，含んでいない場合は，その値を 0 とした．また，標的型攻撃メールであるかどうかを示す二値変数 apt を設け，件名が標的型攻撃メールデータセットに含まれている場合は，その値を 1 と，通常業務メールデータセットに含まれている場合は，その値を 0 とした．このようにして作成した多変量データの概要を**表 2** に示す．

抽出された単語は**図 5** のように単語群 (a)〜(c) に分類された：

(a) 標的型攻撃メールデータセットには出現しない単語であるが，今回抽出した通常業務メールデータセットには出現した単語である．

(b) 標的型攻撃メールデータセットに出現した単語であり，今回抽出した通常業務メールデータセットにも出現した単語である．

(c) 標的型攻撃メールデータセットに出現した単語であり，今回抽出した通常業務メールデータセットには出現しなかった単語である．

本論文における分析手法では，単語群 (b) から標的型攻撃メールの弁別に有効な単語を抽出する．単語群 (a) と (c) は，比較ができないため分析対象外とした．

5.1.2 回帰モデルのあてはめ

次に，単語の出現情報（独立変数）と標的型攻撃メール情報（従属変数）との間に回帰モデルをあてはめ，この回帰モデルにおいて統計学的に有意な独立変数を標的型攻撃メールの弁別に有効な単語とする．

標的型攻撃メールは全数を分析対象としたのに対し，通常業務メールは任意の抽出数を分析対象とした．同様のデータ[*4]を対象に分析するケースコントロール研究では，リスク比は抽出数の影響を受けるため評価に用いることはできず，代わりにオッズ比を用いる [14]．各要因のオッズ比は，ロジスティック回帰分析で求める．

オッズは，事象が起こる（w_i が件名に出現する）確率と起こらない（w_i が件名に出現しない）確率の比である．オッズ比は，一方の群（標的型攻撃メール）におけるオッズと，他方の群（通常業務メール）におけるオッズの比である．オッズ比により，ある単語が件名に含まれない場合と比較して，その単語が件名に含まれる場合に，その件名を持つメールが標的型攻撃メールであるリスクが何倍であるかが分かる．

独立変数として (b) の各単語に対応する変数 $w_1, w_2, \ldots, w_n$ を，従属変数として apt をロジスティック回帰分析に投入し，各単語のオッズ比を推定する．しかし，独立変数は 260 もあり，これらすべてを投入すると回帰分析が収束しないため，変数 $w_1, w_2, \ldots, w_n$ のうち，変

表 2 多変量データ（抜粋）
Table 2 Multivariate data (an extract).

N = 11,455

最頻出 10 単語	品詞	$apt = 1$	$apt = 0$
w_{7089}	名詞–サ変接続	6.67%	0.03%
w_{2815}	名詞–サ変接続	6.53%	0.02%
w_{5749}	名詞–一般	4.16%	0.03%
w_{4159}	名詞–一般	3.22%	0.10%
w_{3423}	名詞–サ変接続	2.98%	0.02%
w_{8403}	名詞–サ変接続	2.83%	0.04%
w_{3357}	名詞–サ変接続	2.73%	0.03%
w_{6489}	名詞–サ変接続	2.68%	0.09%
w_{9035}	名詞–サ変接続	2.57%	0.13%
w_{208}	名詞–サ変接続	2.59%	0.03%

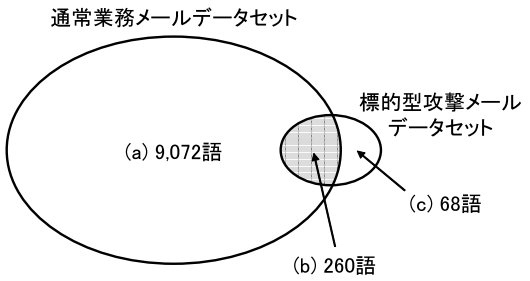

図 5 ユニーク単語数
Fig. 5 Number of unique words.

[*4] 稀な疾病の要因を調査する疫学の研究では，患者データはあるポピュレーショングループの有病者であるのに対し，健常者データは任意の抽出数である．

数 *apt* との相関係数が 0.4 以上[*5]である変数を選びロジスティック回帰分析に投入することにした．$w_1, w_2, \ldots, w_n$ および *apt* は二値変数であるため，相関係数は，テトラコリック相関係数を用いて求めた．ロジスティック回帰分析の結果，**表 3** に示す 42 語のオッズ比が統計学的に有意であり，これらの単語が標的型攻撃メールの弁別に有効であることが示唆された．

5.2 仮説 2 の検証

分析手法 2

本論文では，以下の統計分析手法で，仮説モデルに示した各指標が標的型攻撃の弁別に有効であるかどうかを確認する：

(1) まず，データセットから各指標と標的型メール情報を変数値として持つ多変量データを作成する．

(2) 次に，各指標（独立変数）と標的型攻撃メール情報（従属変数）との間に回帰モデルをあてはめ，この回帰モデルにおいて統計学的に有意な独立変数を標的型攻撃メールの弁別に有効な指標とする．

5.2.1 多変量データの作成

本項では，多変量データの作成について述べる．筆者らは，データセットから**表 4** に示す多変量データを作成した．各変量は，仮説モデルをもとに次のように求めた．

apt 標的型攻撃メールの件名である場合は 1，通常業務メールの件名である場合は 0 の値を持つ変数である．

words 件名における単語数を値として持つ変数である．単語は，5.1.1 項で述べた方法で件名から抽出した．

propernounrate 上記で抽出した単語のうち固有名詞が占める割合を値として持つ変数である．

rankmedian 件名における単語の珍しさの指標を値として持つ変数である．

以下に *rankmedian* の算出方法を述べる．件名は，複数の単語から構成される場合がほとんどである．筆者らは，まず，件名における各単語の珍しさを求め，次に，求めた各単語の珍しさから，件名としての代表値を求め，*rankmedian* の値とすることにした．

まず，各単語の珍しさを求める．筆者らは，単語の珍しさは，その単語の出現頻度に関係し，出現頻度の低い単語は珍しく，出現頻度の高い単語は珍しくないと仮定を置くことにした．各単語の出現頻度は，データセットのすべての件名を含んだ 1 つの文書を想定し，その文書における出現頻度とすることにした．

文書の内容を大まかに把握する目的でキーワードを抽出する研究 [15] や，効率的な英語学習を目的に学問領域や職域などに特有な単語を抽出する研究 [16] があるが，本論文では，監視者の視点から，単語の珍しさを指標化する．

[*5] 人を対象とした社会心理学では，比較的相関があるとされるのは 0.4 以上であることから，相関係数が 0.4 以上の変数を選択した．

表 3 各単語のオッズ比
Table 3 Odds ratio of each word.

N = 11,455

単語	品詞	オッズ比	95%信頼区間	
w_{5252}*	名詞–サ変接続	248.79	61.29	1009.89
w_{3273}*	名詞–サ変接続	248.69	22.35	2767.01
w_{3659}	名詞–一般	135.55	19.66	934.43
w_{8463}*	名詞–サ変接続	127.97	37.42	437.59
w_{607}	名詞–一般	124.35	7.72	2003.29
w_{3142}	名詞–形容動詞語幹	124.35	7.72	2003.29
w_{3994}	名詞–一般	124.35	7.72	2003.29
w_{4660}*	名詞–サ変接続	124.35	7.72	2003.29
w_{5356}	名詞–サ変接続	124.35	17.33	892.35
w_{5454}	名詞–サ変接続	124.35	7.72	2003.29
w_{5502}	名詞–形容動詞語幹	124.35	7.72	2003.29
w_{5762}	名詞–一般	124.35	7.72	2003.29
w_{7339}	名詞–固有名詞–組織	124.35	7.72	2003.29
w_{8970}	名詞–固有名詞–人名	124.35	7.72	2003.29
w_{9028}	名詞–サ変接続	124.35	7.72	2003.29
w_{8628}*	名詞–一般	93.21	16.46	527.93
w_{5007}	名詞–サ変接続	82.90	13.69	502.03
w_{2746}	名詞–サ変接続	69.08	22.71	210.14
w_{7084}	名詞–サ変接続	64.84	2.64	1593.31
w_{535}	副詞–一般	62.17	5.59	691.75
w_{2112}	名詞–一般	62.17	5.59	691.75
w_{3501}	名詞–サ変接続	62.17	5.59	691.75
w_{6391}	名詞–一般	62.17	5.59	691.75
w_{7447}*	名詞–一般	62.17	5.59	691.75
w_{7506}	名詞–サ変接続	62.17	5.59	691.75
w_{9019}	名詞–サ変接続	62.17	5.59	691.75
w_{9113}	名詞–形容動詞語幹	62.17	5.59	691.75
w_{4244}	名詞–形容動詞語幹	49.74	9.53	259.70
w_{3001}	名詞–サ変接続	41.45	4.27	402.23
w_{5310}	名詞–サ変接続	41.45	4.27	402.23
w_{8972}	名詞–サ変接続	41.45	4.27	402.23
w_{4042}	名詞–固有名詞–人名	37.44	1.18	1189.16
w_{911}	名詞–一般	35.53	7.28	173.34
w_{2825}	名詞–サ変接続	31.58	3.56	279.99
w_{5905}	名詞–一般	31.14	1.86	521.56
w_{8837}	名詞–一般	26.57	1.42	498.57
w_{8304}	名詞–固有名詞–一般	25.57	5.33	122.68
w_{6407}	名詞–サ変接続	21.27	1.11	406.71
w_{4158}	名詞–一般	19.16	1.21	303.04
w_{6267}	名詞–一般	18.91	1.63	219.32
w_{8901}	名詞–一般	15.53	1.19	202.47
w_{8456}*	名詞–一般	11.03	1.29	94.15

監視者にヒアリングした内容を考察すると，監視者は，単語を目にする頻度に基づいて判断していると考えられるため，筆者らは，単語の珍しさの算出に出現頻度が適すると判断した．

自然言語は，頻繁に出現する単語はわずかで，めったに出現しない単語はおびただしい数になる傾向を持つ．上位に出現する単語の出現頻度は突出して高いため [17]，頻出

表 4 多変量データの記述統計
Table 4 Descriptive statistics of multivariate data.

N = 9,206

変数	平均値	最小値	最大値
apt	0.02	0	1
words	5.78	1	39
propernounrate	0.08	0	1
rankmedian	97.15	1	144

表 5 単語の珍しさの代表値の計算例
Table 5 Example of calculation for representative values of rareness of words.

件名	単語の珍しさ				平均値	中央値
	w_1	w_2	w_3	w_4		
1	251	254	255	1	190.25	252.50
2	258	253	5	4	130.00	129.00
3	248	9	15	1	68.25	12.00
4	247	2	7	3	64.75	5.00

する単語を注目するためには向いているが，珍しい単語を重視する監視者にとっては，直感的には分かり難い．このため，筆者らは，監視者が直感的にとらえやすいように，出現頻度の順位を求め，これを単語の珍しさの指標とすることにした．この指標の値が小さい場合は，出現頻度が高くありふれた単語であることを意味する．この値が大きい場合は，出現頻度が低く珍しい単語であることを意味する．データセットにおける最頻出単語の珍しさの指標の値は最小値（1）となり，データセットに 1 度しか出現しない単語の珍しさの指標の値は，最大値[*6]となる．

次に，上記のようにして求めた各単語の珍しさから，件名としての代表値を求める．筆者らは，出現頻度順位の平均値と中央値を監視者に提示し，どちらの値が代表値として好ましいか評価することにした．**表 5** に示すように，珍しさが大きく異なる単語が件名に混在する場合，監視者は，件名 1 における w_1，w_2，w_3，件名 3 と 4 における w_2，w_3，w_4（いずれも下線部）のように，多くを占めている珍しさの値を重視していた．単語の珍しさに外れ値を含む件名 1，3，4 は，平均値と中央値に大きな差が生じているが，監視者の感覚に近いのは，中央値であることが分かった．このため，本論文では，各単語の珍しさの中央値を代表値とすることにした．

5.2.2 回帰モデルのあてはめ

単語数や単語出現頻度に関する情報（独立変数）と標的型攻撃メール情報（従属変数）との間に回帰モデルをあてはめ，この回帰モデルにおいて統計学的に有意な独立変数を標的型攻撃メールの弁別に有効な指標とする．

独立変数として *words*，*propernounrate*，*rankmedian* を，従属変数として *apt* を，ロジスティック回帰分析に投入した．推定された各独立変数のオッズ比を**表 6** に示す．*words*，*rankmedian* は，オッズ比の 95%信頼区間が 1 をまたがないため統計学的に有意であり，標的型攻撃メールの弁別に有効な独立変数であった．

ここで投入した独立変数は連続値であるためオッズ比の解釈が 5.1.2 項におけるものと異なる．表 6 に示したオッズ比は，独立変数の値が 1 単位増えたときに標的型攻撃メールのリスクが何倍になるかを示す．独立変数が何単位増えたかをふまえてリスクを解釈する必要がある[*7]．

propernounrate は，オッズ比の 95%信頼区間が 1 をまたぐため統計学的に有意ではなかった．

フィットさせたロジスティック回帰モデルを使用して，*words* を 1 から 20 まで変化させたとき[*8]の標的型攻撃メールの平均予測確率を求めた．その結果を**図 6** に示す．今回の多変量データにおける実際の標的型攻撃メールの確率は 1.88%であるが，多変量データにおけるすべての件名の単語数 *words* が 1 であると仮定した場合には，平均 6.23%の件名が標的型攻撃メールであると予測された．同様に 5 であると仮定した場合は 1.59%，10 であると仮定した場合は 0.27%と予測された[*9]．フィットさせたロジスティック回帰モデルによると，図 6 に示すように，*words* の値が小さ

表 6 各独立変数のオッズ比
Table 6 Odds ratio of each independent variable.

N = 9,206

独立変数	オッズ比	95%信頼区間	
words	0.7013	0.6513	0.7553
propernounrate	0.5131	0.1881	1.3996
rankmedian	1.0078	1.0040	1.0115

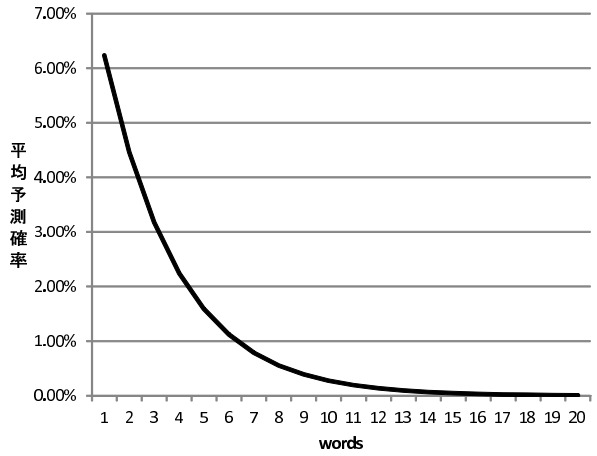

図 6 *words* が変化したときの平均予測確率
Fig. 6 Average predicted probabilities when *words* changed.

[*6] 算出の元となるデータセットによって異なる．本論文で分析対象としたデータセットでは 144 である．

[*7] 独立変数$_i$ の値が n 単位増えたとき，
オッズ比$_i = exp$(回帰係数$_i \times n$) となる．

[*8] 他の変数は平均値に固定．

[*9] 運用において監視対象となるメールにおける標的型攻撃メールの確率は，今回分析に投入した多変量データにおける標的型攻撃メールの確率よりも低いため，本項で示した予測確率は過大評価されていることに注意が必要である．

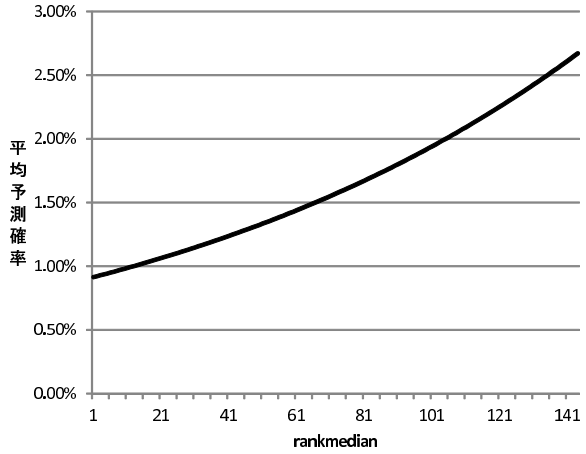

図 7 *rankmedian* が変化したときの平均予測確率

Fig. 7 Average predicted probabilities when *rankmedian* changed.

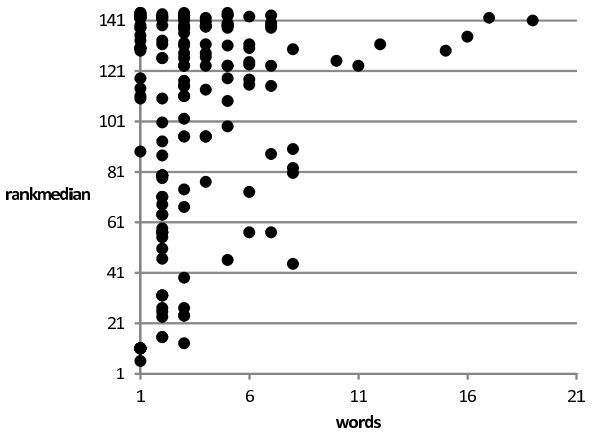

図 8 *words* を横軸に，*rankmedian* を縦軸にとり各標的型攻撃メールをプロットした結果

Fig. 8 Scatter plot of targeted attack mail when X axis is *words* and Y axis is *rankmedian*.

い件名ほど標的型攻撃メールの確率が高く，*words* の値が大きい件名ほど標的型攻撃メールの確率は低くなる傾向があることが示唆された．

フィットさせたロジスティック回帰モデルを使用して，*rankmedian* を 1 から 144 まで変化させたとき[*10]の標的型攻撃メールの平均予測確率を求めた．その結果を**図 7** に示す．*rankmedian* が 1 であると仮定した場合には，平均 0.91%の件名が標的型攻撃メールであると予測された．同様に，72 であると仮定した場合は 1.56%，144 であると仮定した場合は 2.67%と予測された[*11]．フィットさせたロジスティック回帰モデルによると，*rankmedian* の値が大きい件名ほど標的型攻撃メールの確率が高く，*rankmedian* の値が小さい件名ほど標的型攻撃メールの確率は低くなる傾向があることが示唆された．

[*10] 他の変数は平均値に固定．

[*11] 運用において監視対象となるメールにおける標的型攻撃メールの確率は，今回分析に投入した多変量データにおける標的型攻撃メールの確率よりも低いため，本項で示した予測確率は過大評価されていることに注意が必要である．

5.2.3 変数の組み合わせ効果の確認

前項の分析結果により，*words* と *rankmedian* にはそれぞれに主効果があることが示唆された．さらに，たとえば，*words* が小さく，かつ，*rankmedian* が大きい場合に特に多くの標的型攻撃メールの件名が該当するなどの，変数の組み合わせ効果があると，効率良く一次フィルタリングができる実用上のメリットがある．本項では，次のようにして変数の組み合わせ効果の有無を確かめる．

words を横軸に，*rankmedian* を縦軸にとり標的型攻撃メールをプロットし散布図を作成した．その結果，**図 8** に示すようにプロットした点は散布図の左上に集中した．

6. 考察

6.1 分析手法 1 の有効性

表 3 にある 42 単語は統計学的に有意であり，標的型攻撃メールの弁別に有効である単語として基本的には受け入れられる．特に，オッズ比が 100 以上の単語は，95%信頼区間の下限を見ても高いリスクを示しており，特に標的型攻撃メールである危険性の高い単語である．

表 3 の各単語について監視者と検討した結果，「∗」を付した 7 語は，監視者も同様に標的型攻撃メールの危険性が高い単語と認識していることが分かった．

監視者にとって印象の弱い単語であっても，本分析手法により統計学的には危険性が高いとされる単語があった．このため，本分析手法は，監視者に気づきを与え，注意の盲点をカバーしたといえる．

固有名詞について監視者からヒアリングした結果，標的型攻撃メールデータセットに 1 回しか出現しない固有名詞は，統計学的に有意であっても，取り除いても問題ないと解釈してもよいことが分かった．表 3 に 4 つある固有名詞のうち w_{7339}，w_{8970}，w_{4042} は，標的型攻撃メールデータセットに 1 回しか出現せず，監視者は，これらの固有名詞が騙しの細工として繰り返し使われることはないだろうと認識していた．しかし，標的型攻撃メールデータセットに 5 回出現した w_{8304} は，監視者も標的型攻撃メールの危険性が高い単語と認識しているため，このような単語は注意する必要がある．

企業グループ A におけるデータセットを分析手法 1 を用いて分析した結果，標的型攻撃メールの弁別に有効な単語（表 3 のとおり，ただし固有名詞 3 つを除く）を抽出することができた．筆者らは，分析手法 1 は，統計学的に標的型攻撃メールを弁別可能であるので課題 1 を解決できたと考える．

6.2 分析手法 2 の有効性

表 6 を参照すると，*words* が 1 増えたときの標的型攻撃メールであるリスクは 0.7 倍であり，図 6 を参照すると，単

語数が多い件名ほど標的型攻撃メールであるリスクが減少する傾向が示唆された．表 6 を参照すると，*rankmedian* が 1 増えたときの標的型攻撃メールであるリスクは 1.01 倍であり，図 7 を参照すると，珍しい単語で構成される件名ほど標的型攻撃メールであるリスクが増加する傾向が示唆された．

図 8 を参照すると，件名に含まれる単語数が比較的少なく，かつ，件名に含まれる単語の傾向が比較的珍しい標的型攻撃メールが，特に多く，*words* と *rankmedian* の組み合わせ効果があることが示唆された．*words* または *rankmedian* のどちらかに基づきメールを一次フィルタリングするよりも，両者を組み合わせて一次フィルタリングする方が，効率的に詳細調査の対象とするメールを抜き出せるといえる．

ただし，単語数が少なくありふれた単語で構成される件名や，単語数が多く比較的珍しい単語で構成される件名を持つ標的型攻撃メールも少ないながら存在しており，これらの件名を持つ標的型攻撃についても注意しなければならない．

企業グループ A におけるデータセットを分析手法 2 を用いて分析した結果，標的型攻撃メールの弁別に有効な単語の指標が *words* と *rankmedian* であることが明らかになった．筆者らは，分析手法 2 は，統計学的に標的型攻撃メールを弁別可能であるので課題 2 を解決できたと考える．

6.3 メール監視業務への適用に向けた知見

本論文で創出した分析手法 1，2 は，次のようにしてメール監視における一次フィルタリング作業への適用が可能である．あらかじめ，監視者は，組織内のメールデータを採取して，分析手法 1 を用いて，件名に出現すると標的型攻撃メールの危険性が高い要注意単語のリストを作成し，分析手法 2 を用いて，単語数や単語の珍しさの指標の有効性を確認し，閾値を決定しておく．

日々のメール監視では，監視者は，件名に前記の要注意単語リストに載っている単語があるメールや，件名から算出した単語数や単語の珍しさの指標が前記の閾値を超えている（あるいは閾値未満である）メールを機械的に抜き出したのち，残りの差出人と宛先を確認する．

分析手法 1，2 を上記のように運用に適用すると，一次フィルタリング作業における監視者の作業負担を軽減することができる．

さらに，分析手法 1，2 は，検知装置としてシステム化すれば，監視者の手を煩わせることなく，大量のメールの件名を一次フィルタリングにかけるといった運用にも適用可能である．また，分析手法 1，2 は，分析結果に基づいて機械的に抜き出したメールに自動的にラベルを付与して従業員の注意を喚起する運用にも適用可能である．大量のメールを目視確認する際には，この分析手法に基づいて件名を見やすく表示するなどの工夫も可能である．

図 1 とは異なるメール監視運用モデルを採用している組織もある．そのような組織においても，少なくとも，標的型攻撃メールと通常業務メールの両方の記録が残されていれば，これを分析し要注意単語リストと単語数や単語の珍しさの指標の有効性を確認できる．そして，それぞれの組織の運用において，分析で得た要注意単語リストや指標をどのように対策に利用するかを検討する．たとえば小さな組織では，要注意単語などを定期的に従業員に周知し注意喚起するなどの方法も考えられる．

メールの傾向は年月が経つと変化する．また，標的型攻撃メールに施される騙しの細工も巧妙化している．このため最近の傾向を重視する監視者もいる．本論文では，過去 2 年半分のデータに基づいて標的型攻撃メールの弁別に有効な単語を抽出したが，運用においては，複数の異なる期間のデータを対象に定期的に分析を継続する必要がある．

今回考案した分析手法を実運用において評価し，運用に向けた課題を見つけ解決していく必要がある．

6.4 分析手法の限界

分析に投入する標的型攻撃メールの数が非常に少ないことが原因で，表 3 に示した単語のオッズ比の 95%信頼区間が大きくなっている．たとえば，w_{3273} のオッズ比の点推定値は 248.69 であるが，95%信頼区間を見ると，このオッズ比が 22.35 から 2767.01 までの値をとりうることを示す．点推定値を鵜呑みにしないよう注意が必要である．分析に投入する標的型攻撃メールの数を増やすと 95%信頼区間は小さくなり分析結果の確度が高まるが，標的型攻撃メールの特徴から，標的型攻撃が 1 つの組織に大量に送られにくい．このため，この制約は他の組織における分析においてもともなう可能性がある．

監視者にとって印象が強く危険性が高いと認識されている単語であっても，本分析手法により統計学的には弁別に有効ではない単語があることが分かった．これは外部から提供されたり既存の検知装置が添付ファイルの振舞いなどの観点で検知した標的型攻撃メールの件名が，監視者の印象に残っているためと推測される．

今回は分析対象が件名のみであったため，件名のみでは識別不可能な標的型攻撃メールに対応することができない．今後は，本文や差出人など分析対象とする情報を増やしながら分析を継続する．運用においては，本分析手法に基づく識別結果と，既存の検知装置の結果を監視者が見て総合的に判断する必要がある．

7. おわりに

標的型攻撃メールに起因する情報流出事案が繰り返し発生していることが深刻な問題となっている．組織におけるメール監視の現場では，標的型攻撃メールを発見し，配送

を止めたり，受信者の注意を喚起したりして，標的型攻撃メールを使った組織への侵入を防止しようとしている．筆者らは，組織のメール監視業務において，通常業務メールに紛れ込む標的型攻撃メールを見つけようとする監視者の作業を支援することを目的とした．本論文では，メールの件名に施された騙しの細工に着目して標的型攻撃メールを弁別することに焦点を当て，標的型攻撃メールの弁別に有効な単語を抽出する分析手法と，標的型攻撃メールの弁別に有効な単語の指標を明らかにする分析手法を創出した．

これらの分析手法を用いて企業グループ A におけるデータセットを分析した結果，標的型攻撃メールの弁別に有効な単語が明らかになり，また，標的型攻撃メールの弁別に有効な単語の指標がどれであるかが明らかになった．

先行研究 [10] は，対象組織に届いた標的型攻撃メールデータのみを分析した．他方，本研究は，対象組織に届いた標的型攻撃メールデータと通常業務メールデータの両方を分析した．先行研究では，標的型攻撃メールの特徴を記述する分析が可能になった．それに加え，本研究では，通常業務メールデータを分析対象に加えたことにより，ケースコントロール研究とすることができ，通常業務メールに紛れ込む標的型攻撃メールの弁別に有効な単語や指標を，オッズ比を示して明らかにする分析が可能になった．

筆者らは，今後，本論文で創出された分析手法をもとにメールにおける他の要素（件名，差出人，本文，添付文書名など）も分析し，標的型攻撃メールの弁別に役立つより多くの知見を集めたい．本論文の分析対象外である図 5 の単語群 (a) や (c) からもメール監視に役立つ知見が得られると推測している．

本研究により，企業などの組織におけるメール監視運用において監視者の経験を補完して負担を軽減することが期待される．今後は，本研究に基づく対策を様々な組織のメール監視業務に適用し，監視者の負担軽減効果などを検証する必要がある．

本論文における分析手法は，(1) 標的型攻撃メールの問題をかかえており，(2) 情報セキュリティ部門がメールを監視しており，(3) メールシステム運用部門が事例データ提供について協力可能であり，(4) 分析対象期間における標的型攻撃メールと通常業務メールの双方について少なくとも日時と件名の記録が残されている組織に適用できる．今回は，対消費者ではなく対組織の取引が主体で，組織間でのメールのやりとりが多い組織において，有効な分析結果が得られた．このため，筆者らは，同様な他の組織においても有効な分析結果が得られると見込んでいる．

各組織が，本論文と同様の分析を行い，メール監視運用における問題解決に役立てていただきたい．筆者らは，各組織において分析した結果を比較すると，さらに知見が得られると考えている．

参考文献

[1] 警察庁：平成 26 年中のサイバー空間をめぐる脅威の情勢について，警察庁（オンライン），入手先 ⟨http://www.npa.go.jp/kanbou/cybersecurity/H26_jousei.pdf⟩（参照 2016-01-06）．

[2] 日本年金機構：不正アクセスによる情報流出事案に関する調査結果報告について，日本年金機構（オンライン），入手先 ⟨http://www.nenkin.go.jp/oshirase/press/2015/201508/20150820-02.files/press0820.pdf⟩（参照 2016-01-04）．

[3] Mcwhorter, D.: APT1: Exposing one of China's cyber espionage units, Mandiant Corporation（オンライン），入手先 ⟨http://intelreport.mandiant.com/Mandiant_APT1_Report.pdf⟩（参照 2016-02-25）．

[4] Hadnagy, C.: *Social Engineering: The Art of Human Hacking*, Wiley Publishing, Inc., 10475 Crosspoint Boulevard Indianapolis, IN 46256 (2011).

[5] 岡野裕樹, 木邑　実, 辻　宏郷ほか：IPA テクニカルウォッチ「標的型攻撃メールの例と見分け方」，独立行政法人情報処理推進機構技術本部セキュリティセンター（オンライン），入手先 ⟨https://www.ipa.go.jp/security/technicalwatch/20150109.html⟩（参照 2015-12-01）．

[6] 田辺瑠偉，笠間貴弘，吉岡克成ほか：重要情報へのファイルアクセス失敗挙動に基づく情報探索型マルウェア検知手法，情報処理学会論文誌，Vol.57, No.2, pp.597–610 (2016).

[7] 大坪雄平，三村　守，田中英彦：悪性文書ファイル検知のためのファイル構造検査の長期有効性，コンピュータセキュリティシンポジウム 2015 論文集，Vol.2015, No.3, pp.955–962 (2015).

[8] 高橋佑典，渡部正文，島　成佳ほか：おとりの文書を表示する標的型マルウェア検知手法の評価，コンピュータセキュリティシンポジウム 2015 論文集，Vol.2015, No.3, pp.947–954 (2015).

[9] 伊東宏明，青木眞夫：標的型攻撃メールの傾向と事例分析＜2013 年＞，ますます巧妙化，高度化する国内組織への標的型攻撃メールの手口，独立行政法人情報処理推進機構（オンライン），入手先 ⟨https://www.ipa.go.jp/security/technicalwatch/20140130.html⟩（参照 2016-02-14）．

[10] Blond, S.L., Uritesc, A., Gilbert, C. et al.: A Look at Targeted Attacks Through the Lense of an NGO, *23rd USENIX Security Symposium* (*USENIX Security 14*), San Diego, CA, USENIX Association, pp.543–558 (online), available from ⟨https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/le-blond⟩ (2014).

[11] 小森哲志，鍵村達夫：サンプリングとコホート研究，ケース・コントロール研究，薬剤疫学，Vol.18, No.2, pp.95–111（オンライン），DOI: 10.3820/jjpe.18.95 (2014).

[12] Cardo, D.M., Culver, D.H., Ciesielski, C.A. et al.: A Case–Control Study of HIV Seroconversion in Health Care Workers after Percutaneous Exposure, *New England Journal of Medicine*, Vol.337, No.21, pp.1485–1490 (online), DOI: 10.1056/NEJM199711203372101 (1997).

[13] ChaSen—形態素解析器，Nara Institute of Science and Technology（オンライン），入手先 ⟨http://chasen-legacy.osdn.jp/⟩（参照 2016-02-05）．

[14] Fletcher, R.H., Fletcher, S.W. and Fletcher, G.S.: *Clinical Epidemiology: The Essentials*, Wolters Kluwer, Zuidpoolsingel 2, Alphen aan den Rijn, The Netherlands (2012).

[15] 松尾　豊，石塚　満：語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム，人工知能学会論文誌，Vol.17, No.3, pp.217–223（オンライン），DOI: 10.1527/tjsai.17.217 (2002).

[16] 内山将夫, 中條清美, 山本英子ほか：英語教育のための分野

特徴単語の選定尺度の比較，自然言語処理，Vol.11, No.3, pp.165–197（オンライン），DOI: 10.5715/jnlp.11.3_165 (2004).
[17] 関　隆宏，安元裕司，廣川佐千男ほか：教員データにおける高頻度語，情報処理学会研究報告自然言語処理（NL），Vol.2005, No.22 (2004-NL-166), pp.1–8 (2005).

渡部 正文 （正会員）

2002 年奈良先端科学技術大学院大学情報科学研究科博士課程前期修了．同年日本電気株式会社入社．サイバーセキュリティの研究開発に従事．

島 成佳 （正会員）

1997 年北陸先端科学技術大学院大学情報科学研究科博士課程前期修了．1997 年日本電気株式会社入社．2010 年（独）情報処理推進機構に研究員として出向．2012 年電気通信大学大学院情報システム学研究科博士課程後期修了．2013 年 4 月より日本電気株式会社に在籍．サイバーセキュリティの研究開発に従事．博士（工学）．電子情報通信学会会員．

今橋 泰則

1985 年山梨大学工学部卒業．同年 NEC 情報システムズ入社．主に基本システムの開発に従事．近年サイバーセキュリティシステムの運用に携わる．

吉岡 克成 （正会員）

2005 年 3 月横浜国立大学大学院環境情報学府情報メディア環境学専攻博士課程後期修了．博士（工学）．同年 4 月独立行政法人情報通信研究機構で研究員として勤務．2007 年 12 月より横浜国立大学学際プロジェクト研究センター特任教員（助教）．2011 年 4 月横浜国立大学大学院環境情報研究院准教授．マルウェア解析やネットワーク攻撃観測・検知等のネットワークセキュリティの研究に従事．2009 年文部科学大臣表彰・科学技術賞（研究部門），2016 年産学官連携功労者表彰総務大臣賞，受賞．

高木 大資

2012 年東京大学大学院人文社会系研究科博士課程単位取得退学．2015 年より東京大学大学院医学系研究科講師．博士（社会心理学）．社会心理学，犯罪社会学，社会疫学の観点から，地域コミュニティの健康・安全を増進するための研究に従事．日本社会心理学会，日本疫学会，日本公衆衛生学会，地理情報システム学会，犯罪社会学会各会員．