Vol.2019-CSEC-86 No.63
Vol.2019-SPT-34 No.63
2019/7/24

メールにおける誘導手口の推定手法に関する検討

山本 匠[†1]　Bret Harsham[†2]　Ye Wang[†2]　西川 弘毅[†1†3]
上原 航汰[†3]　Chiori Hori[†2]　岩崎 亜衣子[†1]
河内 清人[†1]　西垣 正勝[†3]

概要：本研究では，標的型メールに現れると考えられる，攻撃者の誘導手口を抽出する手法を提案する．誘導手口とは，攻撃者の意図に受信者が心理的に従いやすくなるよう仕向けるテクニック（例えば，権威ある人物を装う，希少性をうったえる）である．機械学習を用いてメール本文中の誘導手口の有無を推定することで，不正メール検知に有効な特徴としての活用や，誘導されていることをユーザに警告するツールへの活用が期待される．

キーワード：標的型メール，誘導手口，チャルディーニの法則，機械学習，ニューラルネットワーク，

1. はじめに

近年，特定の企業や組織を狙った標的型攻撃が増加している．2015 年に起きた日本年金機構への標的型攻撃は記憶に新しい．また，制御システムのネットワーク化に伴い，発電プラントやガスプラントなどの重要インフラへのサイバー攻撃が脅威となりつつあり，国家の安全保障を揺るがす重大な懸念事項となっている．2020 年には世界的に注目の集まる東京オリンピック・パラリンピック競技大会を控えており，攻撃者の恰好のターゲットとなることが予想される．大会期間中にサイバー攻撃により重要インフラが機能停止すれば大会運営に大きな支障が出る．

一方，セキュリティ監視の現場においては，専門的な知識を必要とするスタッフが不足していることが常態化してしまっているのが現状である．経済産業省からの調査報告によると，2016 年時点で 132,060 人の情報セキュリティ人材が不足しており，2020 年には 193,010 人の不足になると予想されている[1]．そのため，少ないスタッフでもサイバー攻撃を高精度かつ効率よく検知することができる技術が必要である．

適切な標的型攻撃対策を実現するためには，標的型攻撃における攻撃者の基本的な行動をサイバーキルチェーンというモデルで整理することが重要と言われている[2]．サイバーキルチェーンでは標的型攻撃を，偵察（Reconnaissance），武器化（Weaponization），配送（Delivery），攻撃（Exploitation），インストール（Installation），遠隔操作（Command and Control），目的実行（Actions on Objective）のステップに分類して整理している．後のステップになればなるほど攻撃が侵攻しており，攻撃を早期に発見し被害を最小限に食い止めるためには，攻撃の初期ステップにあたる標的型攻撃の前段で攻撃を検知することが望まれる．

初期ステップの手口の多くは，対象組織に特化した内容のメールを送る標的型メールと言われている．トレンドマイクロによると，標的型メールによるマルウェア感染は，企業に対する攻撃全体の 76%にも上る[3]．そのため，標的型メール攻撃を防ぐことは，高精度かつ効率よく標的型攻撃を防ぐ観点から重要である．標的型メール攻撃はソーシャルエンジニアリングの典型例であり，攻撃対象者を騙すことで対象者に被害を与える（例えば，情報や金銭を搾取する，PC を不正に操る）．標的型メール攻撃を成功させるためには，標的者に標的型メールを正規のメールと信じ込ませることが不可欠である．

著者らはこれまで，効果的な標的型メールについて，擬態精度および心理操作効力の２つの観点から議論を行い，今後の標的型メールの進化やその対策について検討を行ってきた[4][5][6][7][8]．擬態精度が高い標的型メールとは，標的者にとって正規のメールとの区別がつきにくい標的型メールのことを言う．Open Source Intelligence（OSINT）などを利用し，標的者の所属組織・上司・友人の名前・メールアドレス・出来事・関心事などを取得し，これらの情報をメールに組み入れることによって，擬態精度の高い標的型メール（標的者にとって正規のメールとの区別がつきにくい標的型メール）を標的者ごとに作成することが可能であり[9]，著者らは，OSINT と標的型メール攻撃の相乗効果をモデル化することを試みてきた．具体的には既存研究[4][5]にて，OSINT ツールを利用した擬態精度の高い標的型メールの生成について議論している．本既存研究では，攻撃者が OSINT ツールを用いて攻撃対象の情報を収集していく過程を「状態遷移モデル」として体系化し，その各状態において攻撃者が生成可能な標的型メールを類型化した．

心理操作効力の高い標的型メールとは，人間の行動に対して心理的に影響を与えやすい標的型メールのことである．かねてより人間の行動に対してはある程度の心理操作が可能であることが知られており（チャルディーニの法則[10]），フィッシングメールにもそのテクニックが利用されていることが報告されている[11][12]．またユーザの性格因子に応じて心理操作による誘導の受けやすさが異なることが報告

†1 三菱電機株式会社 Mitsubishi Electric Corporation
†2 Mitsubishi Electric Research Laboratories
†3 静岡大学 Shizuoka University

Vol.2019-CSEC-86 No.63
Vol.2019-SPT-34 No.63
2019/7/24

されていることから[13], 著者らはこれまで, 標的型メールにおける性格と心理操作効力に関する調査を行い，標的型メールにおける心理操作効力には何らかの性格的なものが関係している可能性があることを報告している[6][7][8].

本研究では心理操作効力の高い標的型メールに焦点を当て，メールに使われている心理操作テクニックを特定する方法を検討する．より具体的には，自然言語処理と機械学習の技術を利用し，メール本文中からチャルディーニの法則に該当する表現を抽出する．抽出した情報は，標的型メールを検知するための特徴情報の一つとしての利用や，心理操作効力に影響を受けやすい人物への警告として利用可能である．

2. 従来研究

本章では,既存の不正メール検知技術について紹介する．CipherCraft/Mail[14]は, 受信メールを, 送信ドメイン認証結果や送信経路といった挙動と，名称やアイコン偽装といった添付ファイルに関する不審点をもとに検査し, 自動隔離・注意喚起する技術である．しかし，信頼のおける人物に感染した後に，その人物のメールアドレスを利用してメールを送る攻撃では，挙動に関する不審点は検知できず,高度な攻撃者による添付ファイルが作成される場合，サンドボックスによる検知を通過するため, 本技術では検知できない.

Disarm[15]は，添付ファイルのドキュメントが悪性である可能性があるコード（マクロ等）を含む場合，該当コードを除去し，ドキュメントを再構成することで，悪性マクロの実行を予防する．しかし，マクロ等を活用している組織である場合，Disarm を無効にすることが公式で推奨されているため，そのような組織では有効に働かない.

文献[16]では, 個人ごとに, メール文面に特徴が存在することを利用し，不正なメールを検知する手法を提案している．本手法では，まず不審であるかの識別対象である個人ごとにメールを収集し，個人ごとの特徴量を機械学習アルゴリズムで学習する．学習した分類器により，受信したメールが，予め学習した人物からのものであるかを判定することで，届いたメールが，正しく本人からの文章であるかを判断し, 不審なりすましメールを検知することができる．しかし，認識精度は 67%～100%とまばらであり，確度を持って本人からメールであると言うには信頼性が低いことと，本人識別を通過するように，本人の特徴を学習する巧妙な攻撃には無力である，という課題がある.

文献[17][18]では，メールの対話に着目し，不適切な対話であれば攻撃と判断し検知するソーシャルエンジニアリング対策技術を提案している．自然言語処理を駆使し，メールの対話の中から質問か命令かを特定し，ブラックリストで定義されたソーシャルエンジニアリングの対話手口をもとに不適切な対話かどうかを判断する．ブラックリストは手作業で作成しなければならないという課題がある.

これまで調査したどの既存技術も不正なメールを完全に防ぐことはできない．また単体の検知技術だけでは，巧妙化する標的型メールを完全に検知することは不可能と考える．本研究では，攻撃者の視点に立ち，標的型メールと正規メールの間に差が現れると考えられる心理操作テクニックに着目し，それを特徴として抽出する技術を検討する．この特徴を不正メールの不審な特徴の 1 つとして活用することで，より強固な不正メール検知技術が可能となると期待される.

3. 心理操作効力の高い標的型メール

攻撃者の視点に立つと，標的型メールの本文中には，メールの差出人（攻撃者）がメールの受信者（被害者）に行わせたい何らかの動作が含まれていると考えられる．例えば，添付ファイルをクリックさせたり，URL をクリックさせたり，情報を開示させたりするなどが考えられる．これをメールの「意図」と定義する.

攻撃者は標的型メールに込めた意図を標的者に実行してもらう確率を高くするために，心理的に意図に従いやすくなる文面をメールに含める．これを「誘導手口」と定義する．攻撃に限らず，誘導の手口として良く知られる法則としてチャルディーニの法則がある[10]．チャルディーニの法則は，は，チャルディーニによって提唱された，相手を自分の思い通りに誘導させるための心理法則である．表 1 にチャルディーニの各法則とその概要を記載する.

Akbar らは，メールの本文中にチャルディーニの各法則が組み込まれているか否かを判別するためのフローチャートを開発し，チャルディーニの法則が利用されているフィッシングメールが実際にどの程度存在するのか調査を行った[11]. その結果，Akbar が調査したフィッシングメールデータセット中 96.1%に「権威」の法則が, 41.1%に「希少性」の法則がそれぞれ用いられており，またその他の法則についても高い割合でフィッシングメールに用いられていることが明らかとなった．Akbar の調査は，現在のフィッシングメールにおいてチャルディーニの法則が広く利用されていることを示したが，標的型メールにもチャルディーニの法則が有効に作用することは容易に予想される．Wright らは，被験者である大学生の集団に，チャルディーニの 6 つの法則を利用したフィッシングメールと，チャルディーニの法則を利用していないメールを送り，その反応率（フィッシングメールに書かれている指示に従ってしまう割合）の違いを比較する実験を行った[12]. 実験の結果, チャルディーニのどの法則を利用したフィッシングメールも，利用していないメールと比較して被験者の反応率が高く，フィッシングメールにおけるチャルディーニの法則の効果が確認された結果となった.

チャルディーニの法則そのものは，正常なメールにも利用されることがあるため，この特徴単体では不正なメール

と判定することは難しいだろう．しかしながら，攻撃者の真意を考えると，チャルディーニの法則が不正なメールに利用される頻度は正常なメールに比べると多いのではないかと推測する．メール本文におけるチャルディーニの法則の利用の有無を推定することで，心理操作効力のあるメールに警告を出したり，他の特徴と組合せてより効果的な検知ツールを実現したりすることができると期待する．

4. 提案方式　チャルディーニの法則の推定

4.1 コンセプト

本研究は，心理操作効力の高い標的型メールに利用されると考えられる誘導手口（チャルディーニの法則）の有無をメールの文面から推定することを目的としている．チャルディーニの法則に利用されると考えられるキーワードを一意に定義することは容易ではなく，ルールベースによる特定は難しい．そこで，自然言語処理と機械学習を利用して，チャルディーニの法則に該当すると考えられる文面を推定する技術を提案する．機械学習によりチャルディーニの法則に該当すると考えられるパラグラフを推定するためには，チャルディーニの法則に該当するとラベル付けがされたパラグラフの教師データが必要である．

4.2 データセット

4.2.1 Amazon Mechanical Turk（AMT）**によるデータ収集**

機械学習によりチャルディーニの法則に該当すると考えられるパラグラフを学習するために，教師データとなるデータセットを作成した．教師データの元データとして，Enron Email Dataset[13]を利用した．クラウドソースの Amazon Mechanical Turk（AMT）[14]を利用して，Enron Email Dataset のメールのパラグラフごとチャルディーニの法則の教師データ（ラベル付きのデータセット）を作成した．本来であれば，実際の不正なメールに対してラベル付けし学習データを作成すべきではるものの，不正なメールの実サンプルの入手は非常に困難であるため，公開されている正常なメールのデータセットを利用した．

AMT Worker には，Akbar らのフローチャート[11]参考に，どのような文面の場合にどのチャルディーニの法則に該当するかについて方針を文章で説明し，方針に従いラベル付けのタスクを行ってもらった．1 つのタスクは，1 件のメールに含まれるいくつかのパラグラフに対してどのチャルディーニの法則が該当するかを回答する作業である．1 つのパラグラフに複数のチャルディーニの法則が該当してもよい．AMT Worker には未加工のメールの本文を表示した．返信などの引用文は，可能な限り検出し，背景色を変えラベル付けの対象から外した．1 つのタスクに複数（3 人以上）の Worker を割り当てた．収集したデータセットの情報を表に示す．

表 1　チャルディーニの法則

法則名	概要
権威（Authority）	「肩書や経験などの“権威”を持つ者に対して，信頼を置いてしまう」という心理法則である．自分より立場が上の人物や，目上と感じる人物，特定の分野の専門家には自然と従う心理が生じる．
社会的証明（Consensus）	「周囲の動きに同調したくなる」という心理法則である．「皆がやっているから自分もやる」という気持ちから生じる心理であり，人間は自分以外の誰か（第三者）の行動を物事の判断基準にしてしまう．
一貫性（Consistency）	「自分の行動に一貫性を持たせようとする（持たせたいと考える）」という心理法則である．人間は自ら決めたことに対して，それを正当化する傾向にある．すなわち，過去に経験したような事態に出くわすと，その時と同じ行動を取ろうとする．「表明した約束を守ろうとする」気持ちも，一貫性の法則に含まれる．
好意（Liking）	好意を持っている人からの要請を受けると，積極的に応えようとする」という心理法則である．必ずしも相手のことを知っている必要はなく，「好ましい雰囲気」や「丁寧な口調」等も好意の法則に含まれる．
返報性（Reciprocity）	「人から受けた恩は，返したくなる（返さなければならないと考える）」という心理法則である．一方的に押し付けられた恩であっても返報性が現れる．すなわち，恩を受けた本人が嬉しいか，嬉しくないかに関わらず，何か相手にお返しをしなくてはいけないという心理が働く．
希少性（Scarcity）	「限られたものほど，価値があると感じてしまう」という心理法則である．差し迫った時間的制約があるものや，数が少ないものに対して，それが無くなってしまう前に早く取得しなければならないと思う心理が働く．

Vol.2019-CSEC-86 No.63
Vol.2019-SPT-34 No.63
2019/7/24

表 2　ラベル付けを行った Email の情報

項目	件数
Email 総数	22,988
Email のスレッド総数	10,821
パラグラフ総数	122,625

表 3　ラベル付けされたデータセットの情報

項目	件数
全パラグラフ数	115039
権威（Authority）	6883
社会的証明（Consensus）	1151
一貫性（Consistency）	6166
好意（Liking）	2518
返報性（Reciprocity）	2155
希少性（Scarcity）	1139

4.2.2 一貫性のあるラベル付データセット作成

今回ラベル付けしたデータセットには，複数の異なるメールに短い同じパラグラフ（フレーズや簡単な文章）が含まれていた．例えば，メールの最初と最後の挨拶，署名，定型句，除外漏れの引用文などである．実際のメールにおいても，同様に異なるメールに同じパラグラフが含まれることがある想定し，データセット中の重複は除外せずそのままにした．一貫性のあるラベル付きデータセットを作成するために，データセット中の同一のパラグラフにおいて半数以上の Worker が同一パラグラフに同一のラベルを割り当てた場合に，そのラベルを採用した．

表 3 にラベル付けされたデータセットの情報を示す．

4.3 前処理

パラグラフごと，記号や特殊文字の除去，数字のタグ化などの基本的な前処理を行った

4.4 識別モデル

今回は簡単のため，各チャルディーニの法則に該当するかどうかの 2 値分類の識別器を作成する．

4.4.1 ベースライン方式

比較のために，従来の機械学習アルゴリズムを利用し，チャルディーニの法則の各ラベルを推定するベースラインモデルを作成した．事前実験において，比較的精度が高かった，Logistic Regression を採用した．Google が公開している学習済みの Word2Vec モデル（300 次元）[21]を用いて，パラグラフ中の単語から計算した Word2Vec の平均を特徴情報として利用した．Word2Vec は Google が公開しているモデルをそのまま利用した．Python[22]のライブラリである scikit-learn[23]および gensim[24]を利用し，Logistic Regression による識別器を作成した．

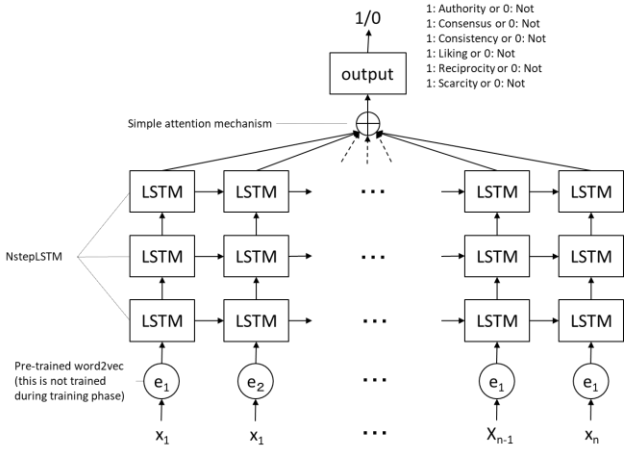

図 1 ニューラルネットワークのアーキテクチャ

Logistic Regression のパラメータである class_weight と C のみグリッドサーチで決定した．それ以外のパラメータはディフォルトのままである．

4.4.2 ニューラルネットワークを利用した方式

図 1 に示すニューラルネットワークを利用し，チャルディーニの法則の各ラベルを推定するモデル（ニューラルネットワークモデル）を作成した．順序関係を学習することができるリカレントニューラルネットワークの 1 つである Long short-term memory（LSTM）を利用しニューラルネットワークを構成した．Google が公開している学習済みの Word2Vec モデル（300 次元）を利用し，embed 層を構成し，パラグラフの単語ごとの分散表現を取得し，一層目の LSTM に入力する．学習時に embed 層は更新されない．

パラグラフ前半の情報についても適切に学習を行うことを目的に，各時間の LSTM の出力に対して注意機構を用いて加重平均をとり出力層に入力する構成とした．陽性データと陰性データの割合をもとに class_weight を調整した。深層学習フレームワークの Chainer を利用してニューラルネットワークモデルのプロトタイプを実装した[25] ．

5. 評価実験

5.1 精度の尺度

チャルディーニの法則のラベルごとに作成した 2 値分類モデルの正検知率（True Positive Rate：TPR）および誤検知率（False Positive Rate：FPR）の関係を表した Receiver Operating Characteristic（ROC）カーブから Area Under Curve（AUC）を算出した．AUC は 1.0 に近ければ近いほど，そのモデルの精度が高いことを意味している．AUC が 0.5 とは，当て推量で 2 択に答えているのと同程度の精度であると判断できる．5-fold cross validation を行い，5

Vol.2019-CSEC-86 No.63
Vol.2019-SPT-34 No.63
2019/7/24

回の検証の AUC とその平均をそれぞれ算出した[a]．5-fold cross validation では，

表 3 に示すデータセットを，ランダムに 5 等分し，3 つを訓練データ，1 つを検証データ，残り 1 つをテストデータとした．

ベースラインモデルについては，各 cross validation において，訓練データと検証データを使ったグリッドサーチによる検証を行い，AUC が最も高いパラメータを決定し，テストを行った．ニューラルネットワークモデルについては，各 cross validation において，epoch ごと訓練データと検証データを用いた検証を行い，AUC が最も高い epoch のモデルをテストに利用した．epoch の数は 20 とした．

5.2 評価結果

ベースラインモデルとニューラルネットワークモデルの評価結果を表 4 と表 5 にそれぞれ示す．表中 AUC は，各 cross validation におけるテストデータによる AUC の値である．また AUC の平均は，5 回分の cross validation の AUC の平均値である．

どちらのモデルに関しても，AUC の平均が 80%を超す結果が得られておらず，十分な精度とは言えないが，機械学習を用いることで，メールのパラグラフ中に存在する，チャルディーニの法則を特定する可能性を示唆することができたと考える．

ベースラインモデルとニューラルネットワークモデルとの比較では，社会的証明と希少性以外の法則において，ニューラルネットワークモデルの方がごくわずかではあるが AUC が高いことが確認された．

6. 考察

ベースラインモデルとニューラルネットワークモデルとの比較では，社会的証明と希少性以外の法則において，ニューラルネットワークモデルの方がごくわずかではあるが AUC が高いことが見て取れる．

表 3 を見ると社会的証明と希少性は他の法則と比べ，陽性データが少ない．ニューラルネットワークにおいては，データ量が精度に強く影響を与えることが知られており，陽性データの少なさがニューラルネットワークモデルの精度が低くなった要因の 1 つと考える．しかしながら，計算量のコストを考えると，ニューラルネットワークを利用する効果があまり得られなかったと考える．

陽性データの数の差に加え，利用する特徴情報やニューラルネットワークのアーキテクチャなどにも，チャルディーニの法則ごと精度に異なる影響が与えられていると考える．例えば，権威（Authority）は，権威にかかわる単語の有無が重要な特徴となりうると考えられる．

表 4　ベースラインモデルによる識別精度

	AUC (CV1)	AUC (CV2)	AUC (CV3)	AUC (CV4)	AUC (CV5)	AUCの平均
権威（Authority）	0.7196	0.7343	0.7292	0.7430	0.7178	0.7288
社会的証明（Consensus）	0.7374	0.7674	0.7723	0.7691	0.7511	0.7595
一貫性（Consistency）	0.6892	0.7007	0.6981	0.6925	0.6959	0.6953
好意（Liking）	0.7215	0.7822	0.7517	0.7670	0.7344	0.7514
返報性（Reciprocity）	0.7551	0.7673	0.7862	0.7868	0.7737	0.7738
希少性（Scarcity）	0.6857	0.7163	0.6958	0.6768	0.6594	0.6868

表 5　ニューラルネットワークによる識別精度

	AUC (CV1)	AUC (CV2)	AUC (CV3)	AUC (CV4)	AUC (CV5)	AUCの平均
権威（Authority）	0.7263	0.7532	0.7541	0.7384	0.7343	0.7413
社会的証明（Consensus）	0.7506	0.7310	0.7844	0.7175	0.7359	0.7439
一貫性（Consistency）	0.6819	0.7036	0.7234	0.6862	0.6878	0.6966
好意（Liking）	0.7395	0.7855	0.7758	0.7512	0.7636	0.7631
返報性（Reciprocity）	0.7666	0.7658	0.8170	0.7683	0.7870	0.7809
希少性（Scarcity）	0.6411	0.6634	0.6896	0.6771	0.6697	0.6682

希少性（Scarcity）は希少性にかかわる情報（例えば，時間，数量などの数）の有無が重要な特徴となりうる．好意（Liking）や返報性（Reciprocity）は，言葉の言い回し，表現の柔らかさ，礼儀正しさなどのフレーズや文章レベルの情報の有無が重要な特徴となりうる．一方，社会的証明（Consensus）や一貫性（Consistency）は，過去に行ったことや他の人の発言や行動など，メールのやりとりなどの文脈が重要な特徴となりうる．ベースラインモデルもニューラルネットワークモデルも，全てのチャルディーニの法則に対して同じもモデルを利用しているため，チャルディーニの法則ごと得て不得手が現れたと考える．今後は，法則ごと適切なモデルについても検討していく予定である．

チャルディーニの法則は正常なメールにおいても使われることがある．そのためこの特徴だけでは不正なメールを特定するまでにはいたらないだろう．しかしながら，攻撃者の真意（攻撃を成功させたい）を考えると，チャルディーニの法則が不正なメールに利用される頻度は正常なメールのそれと比べると多いのではないかと推測する．今後は，正常なメールと不正なメールに含まれるチャルディーニの法則に該当する文章の割合がどの程度異なるのかを，本モデルを利用して検証していく．

7. おわりに

本稿では，標的型メールに現れると考えられる，攻撃者の誘導手口を，メール本文から抽出する手法を提案した．Enron Email Dataset を利用して誘導手口に関するラベル付

a　今回のデータセットは陽性データと陰性データの数の偏りが非常に大きな不均衡データセットである．不均衡データセットにおいて Accuracy（正確度）を精度の尺度に利用することは適切ではない．

データセットを作成し，機械学習による推定を試みた．従来の機械学習を利用したベースラインモデルとニューラルネットワークを利用したモデルの推定精度を比較したところ，ニューラルネッワーク利用した手法の方が若干ではあるが高い精度を示した．どちらのモデルにおいても十分高い精度を示すことができたわけではないが，メール本文中から誘導手口の推定を行う提案方式の実現可能性をある程度示すことができたと考える．

今後はテキストの前処理やニューラルネットワークの改良により，精度改善を行っていく．また実際の不正なメールに対してチャルディーニの法則の推定をした場合に，正常なメールとの間にどの程度の差が現れるかについても調査していく予定である．

参考文献

[1]. 経済産業省, IT 人材の最新動向と将来推計に関する調査結果～ 報告書概要版 ～, http://www.meti.go.jp/policy/it_policy/jinzai/27FY/ITjinzai_report_summary.pdf（2019 年 6 月確認）

[2]. セキュリティサービスグループ編集部，サイバーキルチェーン ～標的型攻撃とアタッカーの活動を知る～ , https://persol-tech-s.co.jp/corporate/security/article.html?id=3（2019 年 6 月確認）

[3]. Trend Micro, COMBATING MALICIOUS EMAIL AND SOCIAL ENGINEERING ATTACK METHODS, https://www.trendmicro.com/cloud-content/us/pdfs/business/datasheets/ds_social-engineering-attack-protection.pdf（2019 年 6 月確認）

[4]. 上原 航汰, 向山 浩平, 藤田 真浩, 西川 弘毅, 山本 匠, 河内 清人, 西垣 正勝, OSINT を利用した標的型メール攻撃手法に関する基礎検討. コンピュータセキュリティシンポジウム

[5]. Kota Uehara, Kohei Mukaiyama, Masahiro Fujita, Hiroki Nishikawa, Takumi Yamamoto, Kiyoto Kawauchi and Masakatsu Nishigaki, “Basic Study on Targeted E-mail Attack Method Using OSINT”, Proceedings of the 33rd International Conference on Advanced Information Networking and Applications (AINA-2019)

[6]. 上原 航汰，井上 佳祐，本多 俊貴，西川 弘毅，山本 匠，河内 清人，西垣 正勝，OSINT と人間の心理を利用した標的型メール攻撃に対するインテリジェンスを活用した防御に関する基礎検討，コンピュータセキュリティシンポジウム 2018（CSS2018）

[7]. 西川 弘毅，上原 航汰，山本 匠，河内 清人，西垣 正勝，インテリジェンスを利用する標的型メールと標的型メールに対するインテリジェンスを利用した防御に関する検討，コンピュータセキュリティシンポジウム 2018（CSS2018）

[8]. 西川 弘毅，山本 匠，上原 航汰，西垣 正勝，河内 清人，標的型メールにおける誘導手口の考察，情報通信システムセキュリティ研究会およびセキュリティ心理学とトラスト研究会（2018 年 3 月）

[9]. Ball LD, Ewan G, Coull NJ. Undermining-social engineering using open source intelligence gathering, in: KDIR 2012: Proceedings of the 4th International Conference on Knowledge Discovery and Information Retrieval, Barcelona, Spain, October 4–7, SciTePress-Science and Technology Publications, 2012.

[10]. Cialdini, R. B. (1987). Influence (Vol. 3). Port Harcourt: A. Michel.

[11]. Akbar, N. (2014). Analysing persuasion principles in phishing emails (Master's thesis, University of Twente).

[12]. Wright, R. T., Jensen, M. L., Thatcher, J. B., Dinger, M., Marett, K. (2014). Research note—influence techniques in phishing attacks: an examination of vulnerability and resistance. Information systems research, 25(2), 385-400.

[13]. Alkış, N., Temizel, T. T. (2015). The impact of individual differences on influence strategies. Personality and Individual Differences, 87, 147-152.

[14]. CipherCraft/Mail, https://www.ntt-tx.co.jp/products/ccraftmailtypeh/（2019 年 6 月確認）

[15]. Disarm, https://support.symantec.com/en_US/article.HOWTO93096.html （2019 年 6 月確認）

[16]. Sevtap Duman, Kubra Kalkan Cakmakciy, Manuel Egelez, William Robertson and Engin Kirda, “EmailProfiler: Spearphishing Filtering with Header and Stylometric Features of Emails”, Computer Software and Applications Conference (COMPSAC), 2016 IEEE 40th Annual.

[17]. Ram Bhakta and Ian G. Harris, Semantic analysis of dialogs to detect social engineering attacks, Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015).

[18]. Yuki Sawa, Ram Bhakta and Ian G. Harris Detection of Social Engineering Attacks Through Natural Language Processing of Conversations, 2016 IEEE Tenth International Conference on Semantic Computing (ICSC)

[19]. Enron Email Dataset, https://www.cs.cmu.edu/~./enron/（2019 年 6 月確認）

[20]. Amazon Mechanical Turk, https://www.mturk.com/（2019 年 6 月確認）

[21]. Google , word2vec , https://code.google.com/archive/p/word2vec/（2019 年 6 月確認）

[22]. Python,

https://www.python.org/downloads/release/python-366/（2019 年 6 月確認）
[23]. scikit-learn Machine Learning in Python, https://scikit-learn.org/stable/（2019 年 6 月確認）
[24]. Genism topic model for humans, https://radimrehurek.com/gensim/（2019 年 6 月確認）
[25]. Preferred Networks, Chainer: A flexible framework for neural networks, https://chainer.org/（2019 年 6 月確認）