G-20

N 階マルコフ連鎖を用いて自動生成した訓練用標的型メールの比較と評価

Comparison and evaluation of automatically generated training targeted emails using Nth-order Markov chains.

○釜田諒大[1]，五味悠一郎[2]
Ryodai Kamada[1], Yuichiro Gomi[2]

Abstract: In this paper, we attempted to solve the problem of automatic generation of targeted emails by using a method of automatic generation based on home-made targeted emails. This method assesses the suitability of auto-generated emails for training, based on homemade targeted emails. We created two types of targeted emails, one using a simple Markov chain and the other using an Nth-order Markov chain, and evaluated them through a blind test by subjects.

1.　まえがき

標的型メールの攻撃件数は平成 30 年に 6740 件，令和元年に 5301 件，令和 2 年に 4119 件と減少傾向にある．攻撃件数が減少することによって被害件数も減少することが予想されるが，被害を断絶するためには日常的にメールを使用する人間がセキュリティ意識や危機感を有する必要がある．セキュリティ意識を向上させる方策としては，企業向けに標的型メールの訓練サービスが提供されている．こうした訓練を依頼する場合は数十万~数百万もの費用となり，現状ではコストがかかり気軽に実施することが難しく，積極的に訓練を行えない問題がある．この問題を解決する方法として，訓練用標的型メールを自動的に生成し訓練に使用する手法を考えた．

2.　問題の性質に対する対処の方針と原理

2.1.　自動生成の手法

マルコフ連鎖は離散的な性質であるため，ランダムな文章が生成できると考え，2019 年の実験ではサンプルとなる標的型メールを自作し，単純マルコフ連鎖で訓練メールの自動生成を行なった．実験の結果，訓練用標的型メールに使用するには相応しいといえない，文脈に不自然さが感じられる文章が生成された．

よって，本報告では N 階マルコフ連鎖は単純マルコフ連鎖よりも文脈が自然な文章が生成されると仮定し，N 階マルコフ連鎖を使用して訓練用標的型メールを生成する手法を考えた．なお，単純マルコフ連鎖との比較を測るため，N 階マルコフ連鎖は 2 階マルコフ連鎖とした．

2.2.　研究目的と目標

企業等の組織が標的型メールの訓練を行う際の障害である費用と手間を軽減し，積極的に訓練を行えるようにするために，訓練用標的型メールを自動生成して配信するシステムを構築し，標的型メール対策としての有効性を示すことを目的とする．本報告では，N 階マルコフ連鎖プログラムから生成される訓練用標的型メールは，精度の高い標的型メールの文章であることをブラインドテストで明らかにし，企業等での標的型メール訓練にも耐えうることを目標とする．

2.3.　作成の元となる標的型メール

IPA では，情報窃取等を目的として，ごく少数または多数ながら特定された範囲のみに対して送られる，利用者の PC をマルウェアに感染させることを目的としたメールを標的型メールと称している．標的型メールの特徴として，受信者に関係がありそうな送信者を詐称していることや，添付ファイルや URL を開かせるため、業務に関係するメールを装い，興味を惹かせる内容にしていることや，添付ファイルの拡張子を偽装する細工などが施されていることが挙げられる．[1]

3.　方法

3.1.　自作標的型メール

自作標的型メールは研究室で作成した 480 通を使用した．作成時に参考にした標的型メールはテーマが異なる合計 12 通で，各メールの最小文字数が 101 文字で最大文字数が 350 文字，平均は約 191 文字で標準偏差は約 71.3 文字だった．自作するメールは，最小文字数と最大文字数および平均や標準偏差がそれぞれ四捨五入した際に同じ値になるようにした．

なお，自作標的型メールの作成時に，態素解析エンジンの Janome を用いて判定された固有名詞(人名，組織名，地域，年号)は伏せ字に置き換えている．

3.2.　訓練用標的型メールの生成

単純マルコフ連鎖では 2019 年の実験で作成した自動生成プログラムを Python3.10.0 で使用した．N 階マルコフ連鎖においては Python3.10.12 を使用し，GoogleColaboratory で自動生成プログラムを作成した．それぞれ，プログラムを実行するたびに 1 通のメールが自動生成される．

１：日大理工・院（前）・情報　２：日大理工・教員・情報

マルコフ連鎖の自動生成手法の場合，自動生成した文章の体裁が整っているとは限らない．よって，本実験では参考にした標的型メールのテーマ 12 種類を元に作成した各テーマ 40 通の自作標的型メールを元に 10 通ずつ自動生成し，生成した 10 通の中から一番体裁の整った 1 通を選択した．単純マルコフ連鎖プログラムと N 階マルコフ連鎖プログラムの両方で生成し，結果として 12 通×2 の 24 通をブラインドテストに使用した．

3.3.　訓練用標的型メールの精度評価

自動生成した文章精度の高さをブラインドテストで評価した．

ブラインドテストでは，自動生成した訓練用標的型メールの文章が「人が作成した標的型メール」か「プログラムが生成した標的型メール」かを被験者に選択してもらった．評価結果の「人が作成した標的型メール」が回答割合の 5 割かそれ以上を得られれば，文章精度が高くランダム性も含んだ標的型メールを自動生成できたといえる．

各プログラムが生成した標的型メールの比較を行うため，ブラインドテストの問題は，前半の 12 問は単純マルコフ連鎖プログラム，後半の 12 問は N 階マルコフ連鎖プログラムの計 24 問とした．被験者は日常的にメールを使用している 20 代から 60 代の社会人 22 人と 20 代の学生 16 人の計 38 人とした．

4.　結果と考察

ブラインドテストの結果を Figure 1,2 に示す．

実験の結果，単純マルコフ連鎖を用いて自動生成したメールが「人が作成した標的型メール」と判断された割合が 5 割以上のメールは 12 通中 3 通であった．対して，N 階マルコフ連鎖を用いて自動生成したメールが「人が作成した標的型メール」と判断された割合が 5 割以上のメールは 12 通中 7 通であった．この結果から，単純マルコフ連鎖を用いて生成した文章よりも，N 階マルコフ連鎖を用いて生成した文章のほうが，体制の整っている自然な文章であると考えられる．

「人が作成した標的型メール」と判断された割合が 5 割以上のメールは，自己紹介をしており，受信者へ求められていることが率直に読み手に伝わる文章であるという共通点があった．よって 5 割以上と判断されたメールは，訓練用標的型メールに使用できると考えられる．

メールの自動生成には，2019 年の実験で作成した単純マルコフ連鎖プログラムと，自作した N 階マルコフ連鎖プログラムを使用した．どちらもマルコフ連鎖を用いたプログラムだがソースコードの書き方に差異があるため，文章生成に影響する可能性がある．この改善策として，自作した N 階マルコフ連鎖プログラムを N=1 に設定することで，より正確に比較できると考えられる．

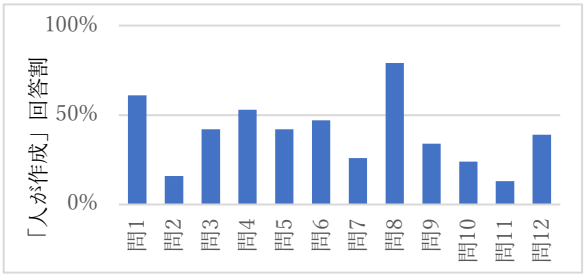

Figure 1 Percentage of respondents who answered that the targeted e-mail was created by a Human [%].-simple Markov chain.

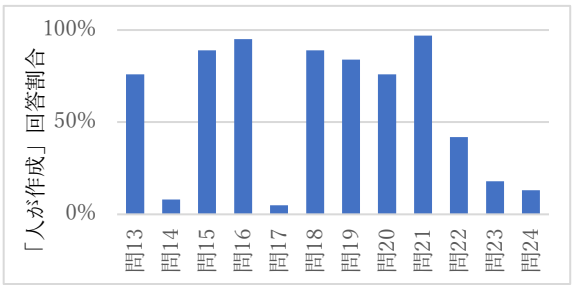

Figure 2 Percentage of respondents who answered that the targeted e-mail was created by a Human [%].-Nth-order Markov chain.

5.　まとめ

ブラインドテストの結果，自動生成した訓練用標的型メールは，N 階マルコフ連鎖の方が単純マルコフ連鎖よりも，標的型メール訓練に使用するメールとして相応しいことがわかった．しかし，生成したすべてのメールが訓練用として相応しいとはならなかった．

訓練用として相応しくないと判断された文章を解決する方法としては，別の自動生成手法を使用することや，元となる自作標的型メールのサンプル数をさらに増やすことが考えられる。こうした改善により，企業等での訓練にも耐えうる，精度の高い標的型メールの文章が生成されると期待できる．

6.　参考文献

[1] J-CRAT／標的型サイバー攻撃特別相談窓口：https://www.ipa.go.jp/security/tokubetsu/index.html，[アクセス日 2022-1-18].