

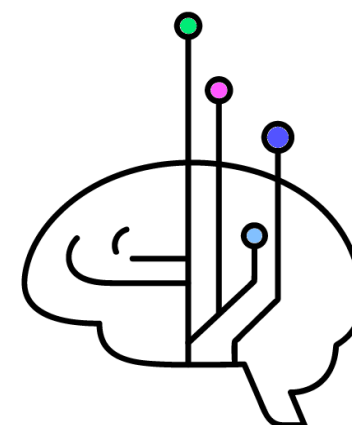
令和3年度 卒業研究発表会

迷惑メール解析における AI 技術活用の検討

山形県立産業技術短期大学校庄内校 情報通信システム科

【発表者】[リーダー] 齋藤 力、[サブリーダー] 制野 隼人、荒川 楓麻、五十嵐 鼓太、池田 陸

【指導教員】開沼 和広、工藤 智子



AisisTechnology



説明内容

1. 背景
2. 研究の目標
3. 現状調査
4. AIへの活用
5. システムの実装
6. 結論
7. 今後の展望
8. 参考文献



1 背景

1.1 国内の迷惑メール状況

- 情報セキュリティ10大脅威 -

迷惑メールによる脅威が上位に位置している

表1.1 情報セキュリティ10大脅威 2021(抜粋)

順位	個人	組織
1位	スマホ決済の不正利用	ランサムウェアによる被害
2位	フィッシングによる個人情報等の詐取	標的型攻撃による機密情報の窃取
3位	ネット上の誹謗・中傷・デマ	テレワーク等のニューノーマルな働き方を狙った攻撃
4位	メールやSMS等を使った脅迫・詐欺の手口による金銭要求	サプライチェーンの弱点を悪用した攻撃
5位	クレジットカード情報の不正利用	ビジネスメール詐欺による金銭被害

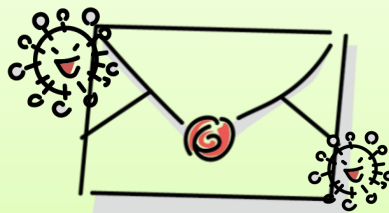
情報セキュリティインシデントの原因



1.2 本校の迷惑メール状況

本校には過去4年間で**約2.5万件の迷惑メール**が保存されている

- 現在の迷惑メール件数 -



迷惑メール

24,952件

フィルタに引っかかった迷惑メール

活用

本校専用の「**AIによる迷惑メール判定システム**」を構築したい



1.3 迷惑メールとは

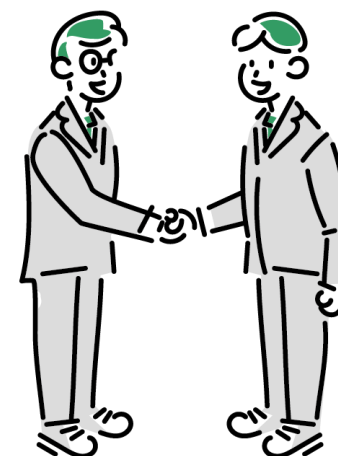
－ 迷惑メール －

「受信者の意向を無視して一方的に繰り返し送り付けられるメール」

「怪しいWebサイトへの誘導や、

広告宣伝の勧誘メールなどを無差別に送信するメール」

- 詐欺・なりすまし
- 架空請求
- 副業勧誘
- 広告宣伝
- 出会い系 など



2.研究の目標



- 1. 現状調査 -

本校に保存されている迷惑メールの特徴を分析し、
国内の迷惑メールの動向と比べ、庄内校独自の特征があるかを解析する

- 2. システムへの実装 -

本校のデータを学習データとして適応するAI技術の比較検討を行い、
迷惑メール判定の実装を検討する



3. 現状調査

3.1 本校に届いた迷惑メールデータの概要

- 本校のメールシステム -

本校のメールシステムのフィルタには、procmailを使用している

- procmail -

受信メールをフィルタにかけて様々なディレクトリに分類し、スパムメッセージとして除外できるソフトウェア

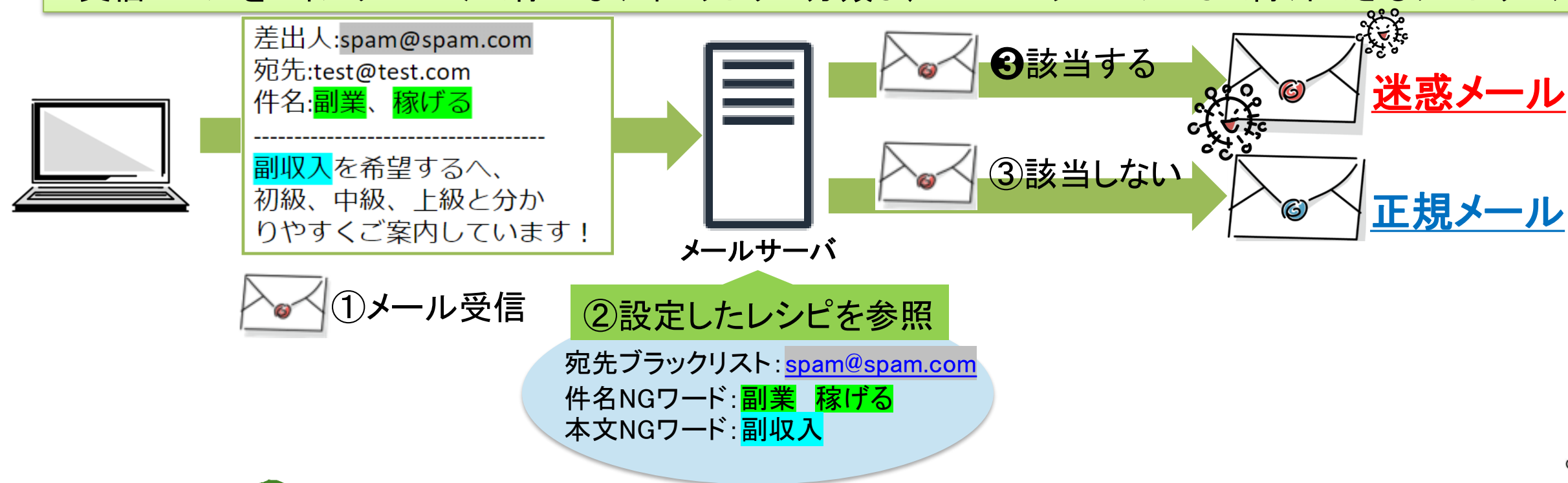


図3.1 procmail



3.1 本校に届いた迷惑メールデータの概要

- フィルタが検知した迷惑メール -

概要: 本校が受信したメールのうち、迷惑メールと判定されたもの

データ取得期間: 2017年10月～2021年12月

データ件数: 24,952件



②設定したレシピを参照

宛先ブラックリスト: spam@spam.com

件名NGワード: 副業 稼げる

本文NGワード: 副収入

図3.2 迷惑メール



3.1 本校に届いた迷惑メールデータの概要

- フィルタをすり抜けた迷惑メール -

概要: 担当教員が受信したメールのうち、正規メールと判定された迷惑メール
データ件数: 数十件

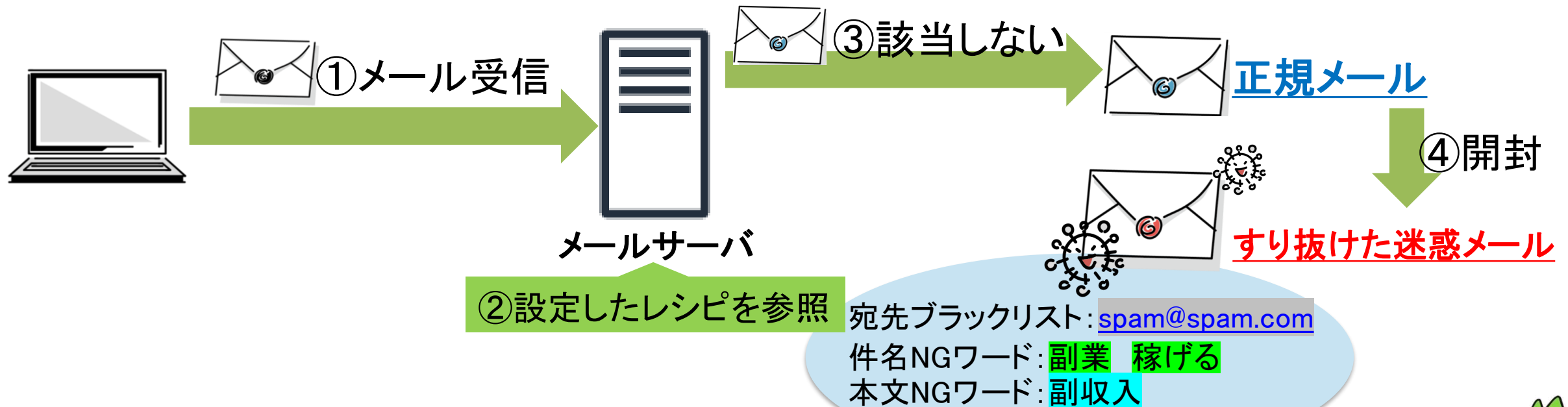


図3.3 正規メール

3.2 本校に届いた迷惑メールデータの調査

- 調査方法 -

本校が受信したメール(24,952件)の約10%を用いて調査を行う

【迷惑メール】項目別総数

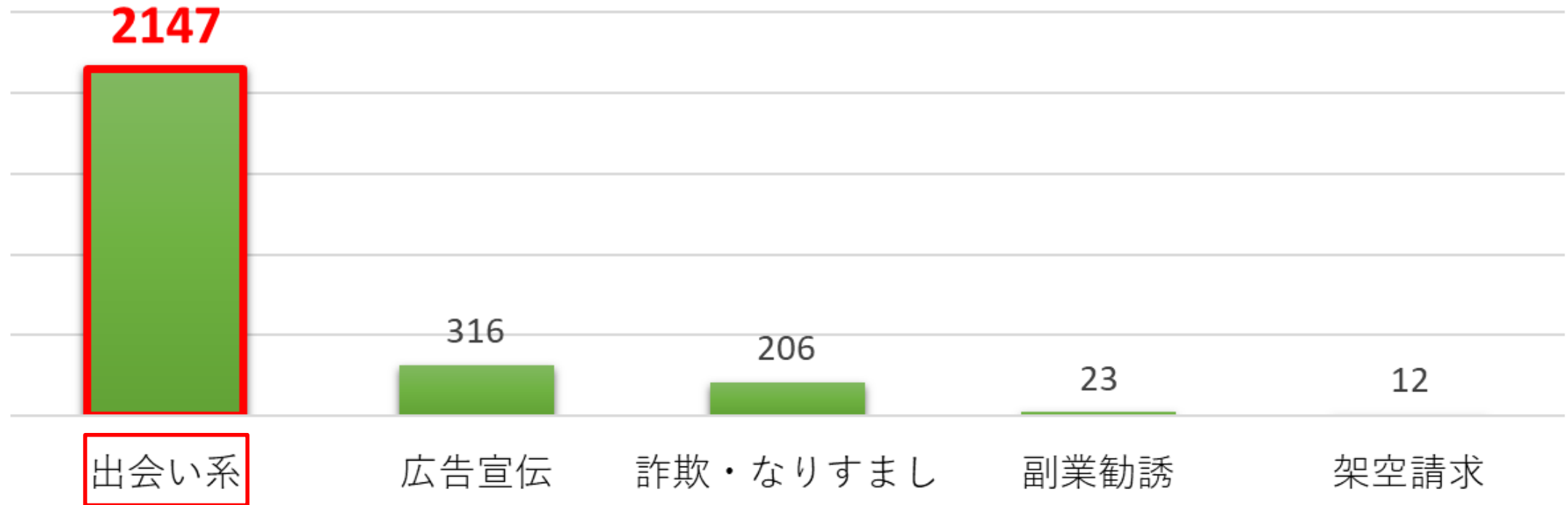


図3.4 項目別総数

3.2 本校に届いた迷惑メールデータの調査

【迷惑メール】項目別割合

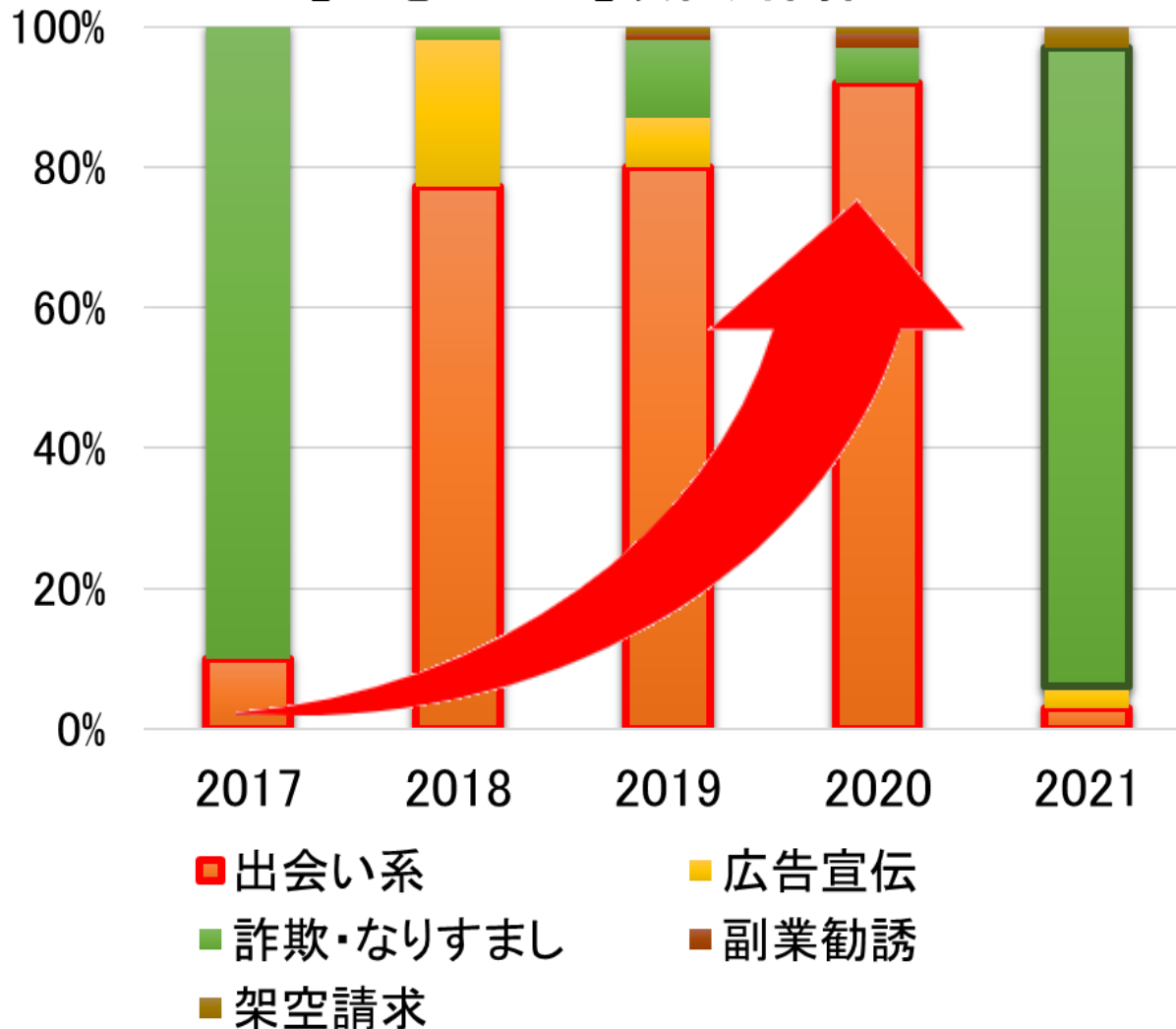


図3.5 項目別割合

- 2020年まで**出会い系**の割合が増加

- 近年は、大手カード会社を名乗る**詐欺・なりすましメール**が増加



4. AIへの活用

4.1 本研究の迷惑メール判定に使用を検討するAI技術

ナイーブベイズ
(ベイジアンフィルタ)

- ベイズ推定と呼ばれる統計的手法を応用した有名な機械学習手法

検討理由

迷惑メール対策ソフト

- SpamAssassin
- POPFile
- BkBsfilWrapper など、**多くの迷惑メールソフトで採用**されている

特徴

- ベイジアンフィルタで利用されているアルゴリズム
- あるデータがどのカテゴリに属するかを確率的に求める
- シンプルなアルゴリズム
- 大きなデータセットに対しても有効



4.1 本研究の迷惑メール判定に使用を検討するAI技術



決定木

- 段階的にデータを分割していき、木のような分析結果を出力する

検討理由

- 先行研究を調査したところ使用されていた
- **テキスト分類**に使用されるアルゴリズム

[2] SPAMメールの判別に適した機械学習

特徴

- 解釈が容易である
- 過剰適合してしまいがち

↓ 対策

ランダムフォレストなど

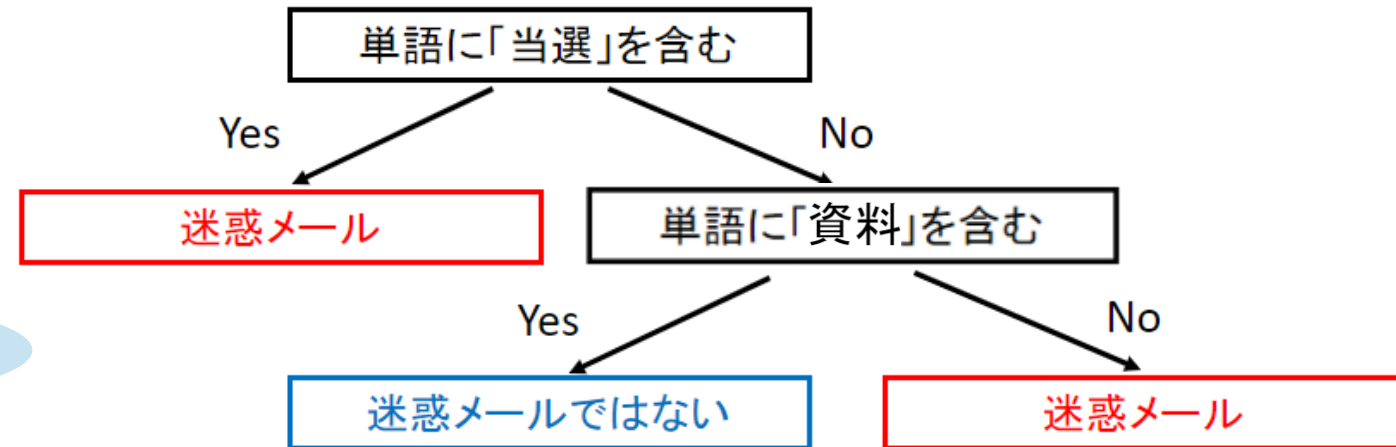


図4.1 決定木



4.1 本研究の迷惑メール判定に使用を検討するAI技術

One-Class SVM (異常検知)

- 過去のデータを正常値として新たなデータが特異か否かを判定する手法

検討理由

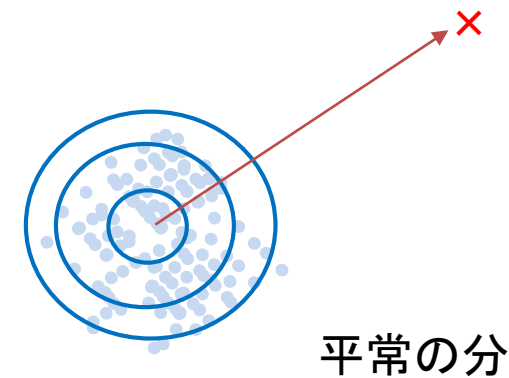
- 迷惑メールのみ**で学習モデルを作成できる

特徴

- 教師なし学習
(入力したデータのみが学習データとして与えられる)
- 正常データのみで学習モデルを作成

変数2

異常なデータ
分布とのギャップが大きい(特異な値)



平常の分布

変数1

図4.2 異常検知

4.2 本研究で使用するソフトウェア

データ分析ソフト

– Tableau –



- 膨大なデータに対応できる

統合開発環境

– Jupyter Notebook –



- Webブラウザ上でPythonが記述・実行できる統合開発環境

クラウドサービス

– Microsoft Azure Machine Learning Studio –



- 機械学習、データ分析などをクラウド上で行うためのプラットフォームを提供



4.3 学習データの前処理

決定木・ナイーブベイズ

「判定用タグ」、「本文」から構成されるCSVファイルを作成する

```
spam,Fuckbuddy Alayah Alba ne  
spam,新型コロナ せい お金 時間  
spam,Cripto List illion Coinb  
spam,Eager Hookup Aaliyah Abb  
spam,クレジットカード 情報 更  
spam,ホット 女の子 女性 今夜  
spam,お客様 各位 2021年 11月3  
spam,Alanis Eager Hookup Alex
```



図4.3 【決定木・ナイーブベイズ】前処理済csvファイル

4.3 学習データの前処理

One-Class SVM(異常検知)

メール

	24時間	48時間	abi	abia	ac
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	10

単語

メールに登場する単語の出現頻度

図4.4 【異常検知】前処理済csvファイル



4.3 学習データの前処理

- 自動化範囲 -

1. 本文のデコード
2. HTML文からテキストの抽出
3. URLなどの不要な文字の削除
4. 形態素解析
5. 単語の出現回数を数える

- 手作業範囲 -

1. 使用データの選択
2. デコードの実行
3. csv入力

```
spam,Fuckbuddy Alayah Alba ne  
spam,新型コロナ せい お金 時間  
spam,Cripto List illion Coinb  
spam,Eager Hookup Aaliyah Abb  
spam,クレジットカード 情報 更  
spam,ホット 女の子 女性 今夜  
spam,お客様 各位 2021年 11月3  
spam,Alanis Eager Hookup Alex
```

図4.3【決定木・ナイーブベイズ】前処理済csvファイル

単語の出現回数を数えて、csvファイルにする

	24時間	48時間	abi	abia	ac
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	10

図4.4【異常検知】前処理済csvファイル



4.4 AI技術の評価

各AIのスコア比較

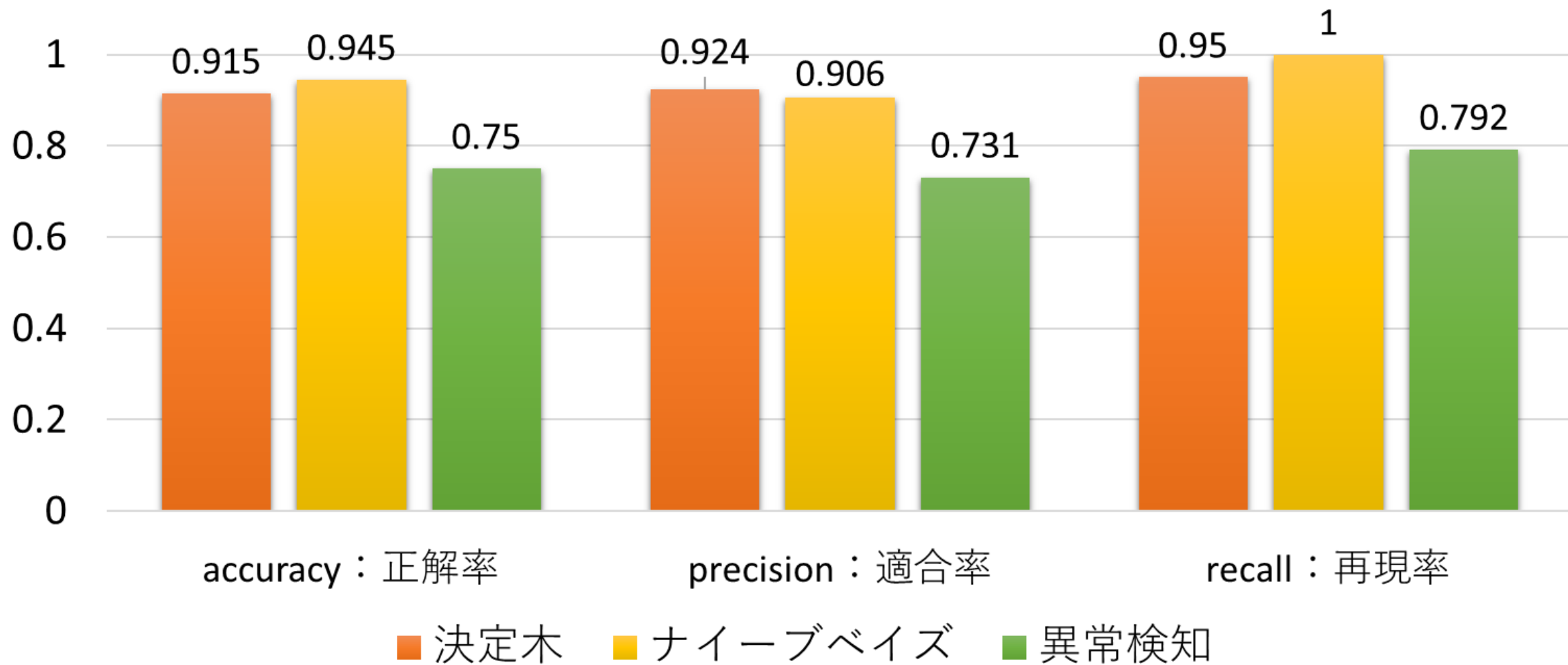


図4.5 AI技術の比較



4.4 AI技術の評価

正解率
(accuracy)

- 正解を当てることのできた割合

適合率
(precision)

- 高いほど正規メールを誤って迷惑メールと判定しなくなる
- **誤検出を少なくしたい場合**に重要視

再現率
(recall)

- 高いほど迷惑メールを誤って正規メールと判定しなくなる
- **取りこぼしを少なくしたい場合**に重要視

ユーザーが正規メールを見逃さないようにしたい...



適合率が高いAI技術を採用

4.4 AI技術の評価

各AIのスコア比較

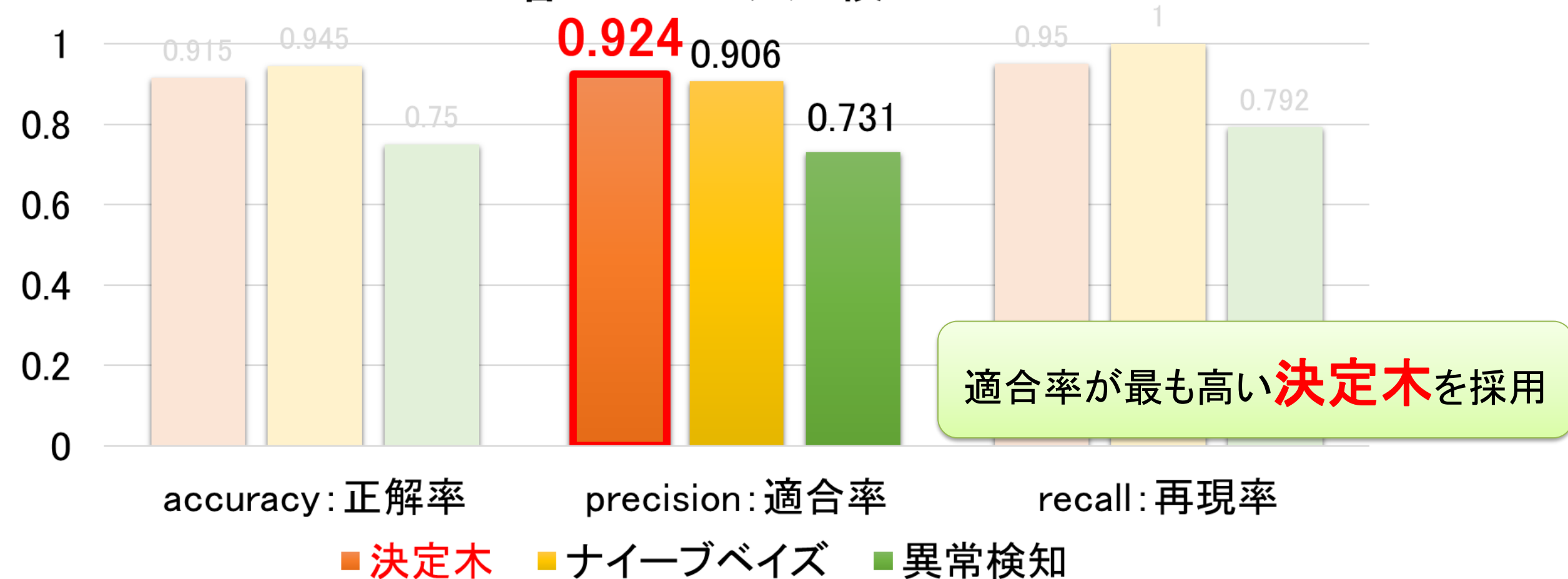


図4.5 AI技術の比較

4.5 決定木の考察

－ 手順 －

1. 各年、ランダムで抽出した迷惑メールと共通の正規メール各100件を教師データとする
2. 判定を行いたいメールを受信した年の教師データをテストデータとし、
その前の年の教師データを訓練データとする
3. 2.で作成したモデルで判定を行い、結果を確認する

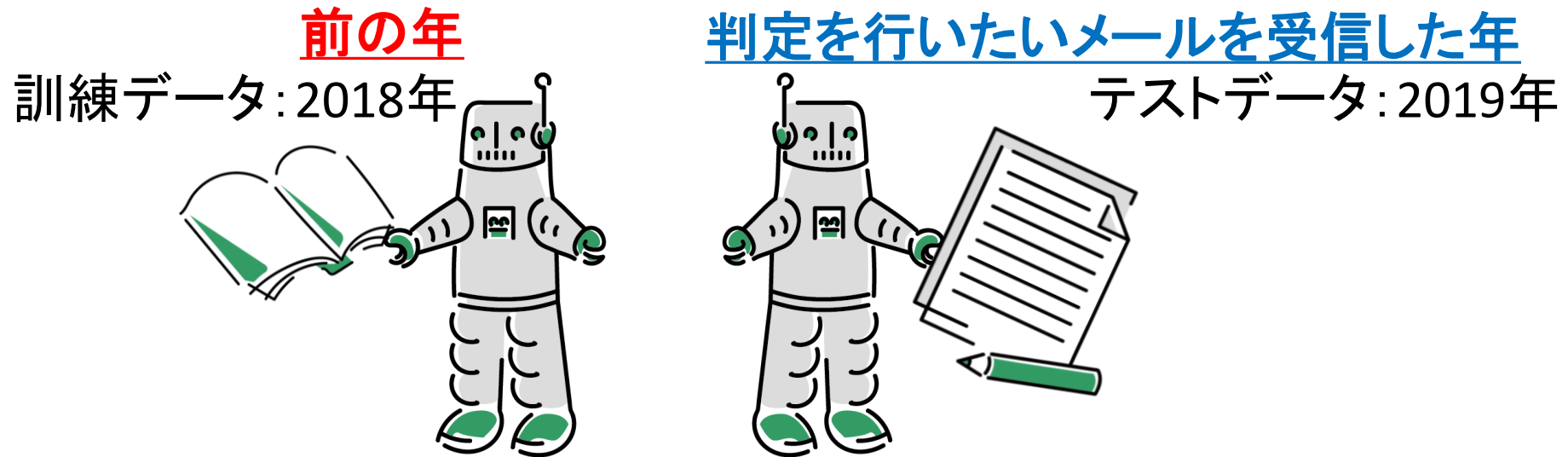


図4.6 手順の例

4.5 決定木の考察

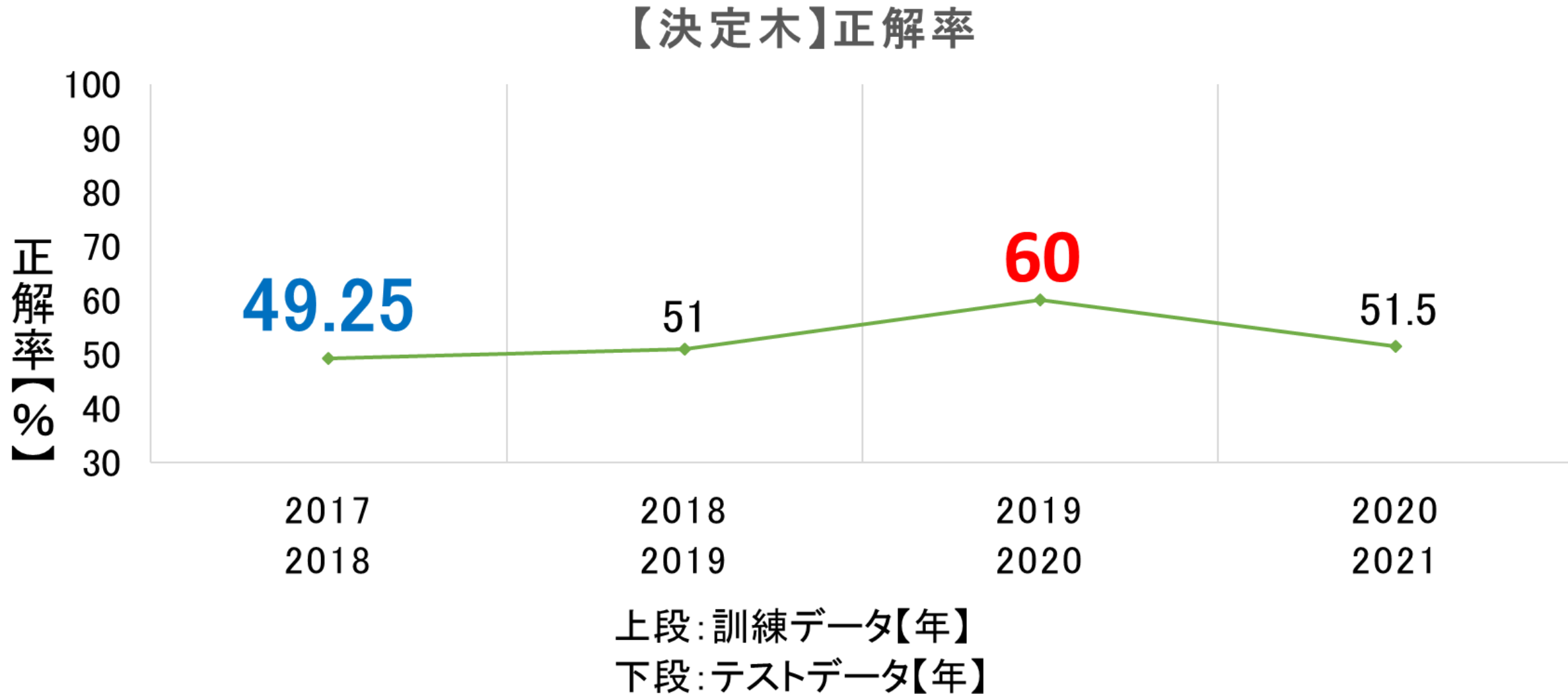


図4.7 決定木の正解率



4.5 決定木の考察1

表4.1【訓練データ:2018・テストデータ:2019】寄与度ランキングと出現回数

(個)	spam	ham
1. メール	0	570
2. unsubscribe	40	4
3. mail	85	50
4. こちら	0	418
5. よう	0	284
6. 問い合わせ	0	260
7. email	82	14
8. want	104	2
9. photos	47	0
10. send	31	2

spam、hamどちらのメールにも
特徴となる単語が存在している



悪用すると、

正規メールらしい迷惑メールを送ること

が可能になってしまう



4.5 決定木の考察2

同じメールが**大量**



- **機械的**に送られてきている
- 組織宛ての標的型メールではなく、
個人宛の迷惑メールである



25	You seem like my type and I would like to know you more! Write me if you
26	You seem like my type and I would like to know you more! Write me if you
27	You seem like my type and I would like to know you more! Write me if you
28	You seem like my type and I would like to know you more! Write me if you
29	You seem like my type and I would like to know you more! Write me if you are inter
30	You seem like my type and I would like to know you more! Write me if you
31	You seem like my type and I would like to know you more! Write me if you
32	You seem like my type and I would like to know you more! Write me if you
33	You seem like my type and I would like to know you more! Write me if you
34	You seem like my type and I would like to know you more! Write me if you
35	You seem like my type and I would like to know you more! Write me if you
36	You seem like my type and I would like to know you more! Write me if you
37	10 reasons to try absolutely new enhancing product. You don't have to spe
38	You seem like my type and I would like to know you more! Write me if you
39	You seem like my type and I would like to know you more! Write me if you

図4.8 ランダムに抽出した迷惑メール

4.5 決定木の考察3

8	Virtual Pilot 3D 2019 Official Flight
9	Virtual Pilot 3D 2019 Official Flight
10	Looking for hot girls and milfs? Want
11	Virtual Pilot 3D 2019 Official Flight
12	Virtual Pilot 3D 2019 Official Flight
13	Looking for hot girls and milfs? Want
14	Looking for hot girls and milfs? Want
15	This is something you should NEVER mis
16	With Step By Step Instructions Softwar

図4.9 【2019年】ランダムに抽出した学習データ

8	Greetings, My name is Felix, a Chief
9	ホットな女の子と女性をお探しですか？ 今夜
10	ホットな女の子と女性をお探しですか？ 今夜
11	Greetings, My name is Felix, a Chief
12	ホットな女の子と女性をお探しですか？ 今夜
13	ホットな女の子と女性をお探しですか？ 今夜
14	ホットな女の子と女性をお探しですか？ 今夜
15	ホットな女の子と女性をお探しですか？ 今夜
16	ホットな女の子と女性をお探しですか？ 今夜

図4.10 【2020年】ランダムに抽出した学習データ



年によってメールの傾向が違うため、
前年のデータを用いて学習を行うと結果が悪くなってしまう



定期的に学習させ続けることが必要になる



5 システムの実装【概要】

- システム実装手順 -

- ① 判定用プログラムをメールサーバに組み込む
- ② 受信と同時に迷惑メール判定を行い、迷惑メールだった場合ユーザに通知する

使用するAI技術：決定木

- メールサーバ概要 -

OS	CentOS 6.4
MTA (メール転送エージェント)	Postfix 2.6.6
メール格納形式	Maildir形式
Webサーバソフトウェア	Apache 2.2.15

5.1 システムの実装【概要】

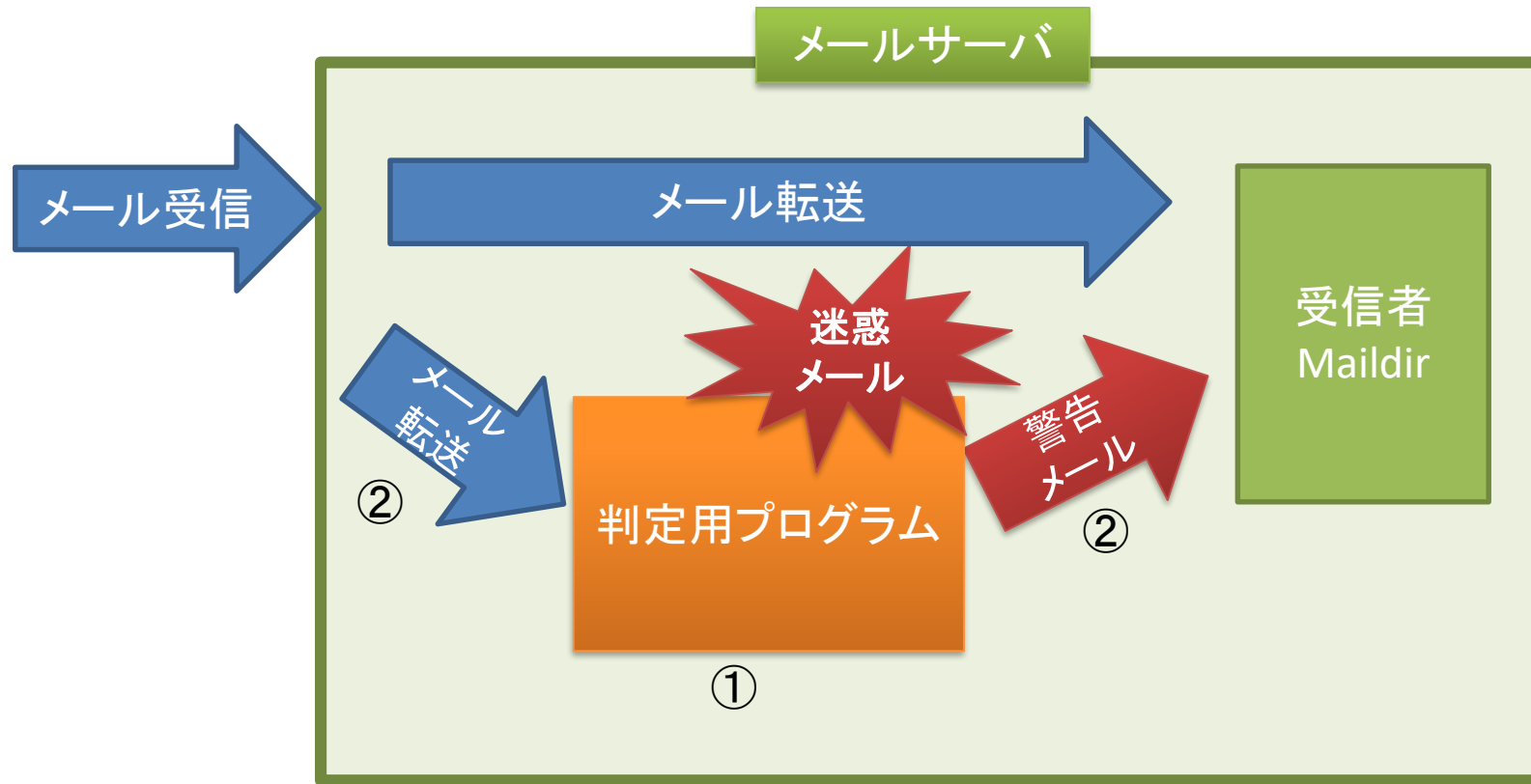


図5.1 システム概要図

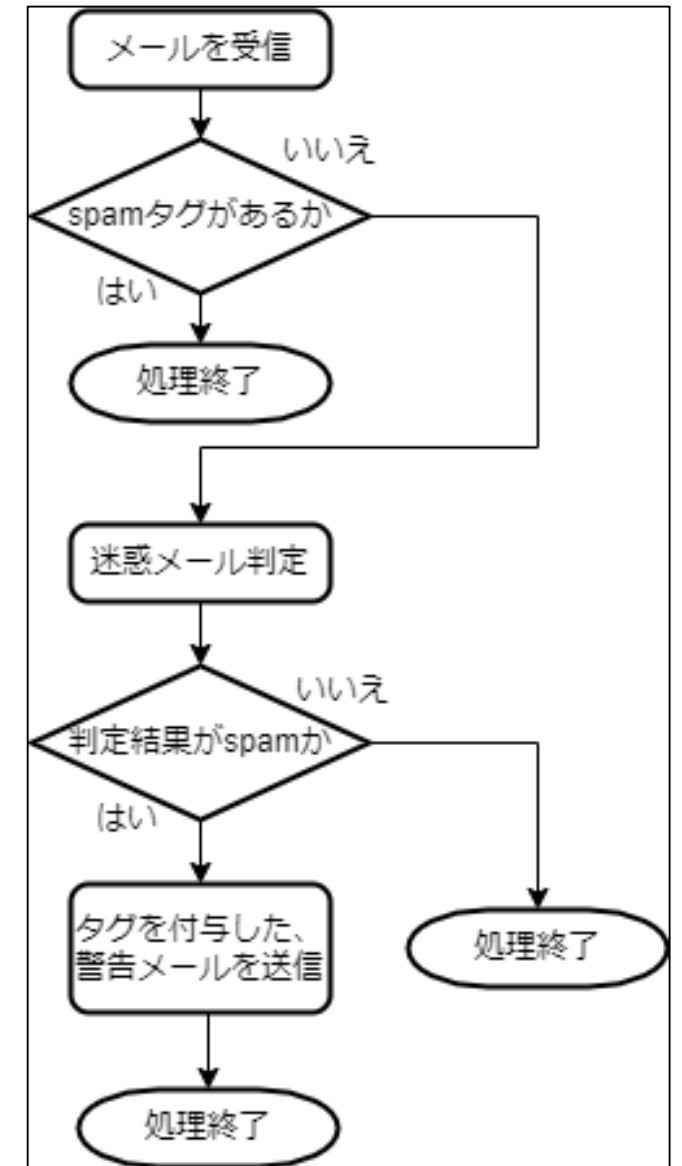


図5.2 プログラムフローチャート



5.2 システムの実装【実行画面】

- 受信フォルダ -

現在のフォルダ: 受信箱

[ログアウト](#)

[メッセージ作成](#) [アドレス帳](#) [フォルダ](#) [オプション](#) [検索](#) [ヘルプ](#) [自動返信](#)

[全反転](#)

1 - 2 件目を表示中 (2 件中)

チェックしたものの移動:

チェックしたものの状態変更:

差出人	日付	件名
☐ Nobody	1:54 pm	[*spam*] Urgent Request Concerning Your Invoice
☐ [REDACTED] ac.jp	1:54 pm	Urgent Request Concerning Your Invoice

[全反転](#)

スパム警告メール

受信したメール

図5.3 受信フォルダ

件名に**spamタグ**が付与される

5.2 システムの実装【実行画面】

- スпам警告メールの内容 -

受信した迷惑メールの
ソースデータ

件名: [*spam*] Urgent Request Concerning Your Invoice
差出人: "Nobody" <[redacted]@cit.ac.jp>
日付: 2022年 2月 10日 (木) 1:54 pm
宛先: [redacted]@cit.ac.jp
重要度: 中
オプション: [ヘッダの表示](#) | [印刷用バージョン](#) | [このファイルをダウンロード](#)

```
Return-Path: [redacted]@cit.ac.jp>
X-Original-To: [redacted]@cit.ac.jp
Delivered-To: [redacted]@cit.ac.jp
Received: from [redacted] (localhost [127.0.0.1])
  by [redacted] (Postfix) with ESMTP id AE22C80BD3
  for [redacted]@cit.ac.jp; Thu, 10 Feb 2022 13:54:12 +0900 (JST)
Received: from 172.[redacted]
  (SquirrelMail authenticated user [redacted]
  by [redacted]@cit.ac.jp with HTTP;
  Thu, 10 Feb 2022 13:54:12 +0900
Message-ID: <3675ba50218e44f2786c810ee31f685.squirrelmail@[redacted]@cit.ac.jp>
Date: Thu, 10 Feb 2022 13:54:12 +0900
Subject: Urgent Request Concerning Your Invoice
From: [redacted]@cit.ac.jp
To: [redacted]@cit.ac.jp
User-Agent: SquirrelMail/1.4.22
MIME-Version: 1.0
Content-Type: text/plain; charset=iso-2022-jp
Content-Transfer-Encoding: 8bit
X-Priority: 3 (Normal)
Importance: Normal
```

```
My name is Taro Tanaka and I work at Japan.
We are happy to inform you that we received your payment.
We are sorry to inform you that the item you ordered is out of stock.
Please contact us at your earliest convenience.
I look forward to hearing from you.
```

ヘッダ

本文

図5.4 スпам警告メール



5.3 システムの実装【結果】

- 実現した機能 -

- メール受信をトリガーとしたプログラムの実行
- 受信メールの迷惑メール判定
- 迷惑メール検知時の通知

- まとめ -

- 迷惑メール判定システムとして、**最低限の機能を実装**できた
- **新しい学習データや別の機械学習アルゴリズム**がより最適と分かった場合、**容易に変更可能な状態**

6. 今後の展望

□ 判定精度の向上

- 今後の増加分も含め、すべての蓄積データを使用した分析・システム実装
- 再学習機能の実装
- 様々な言語への対応

□ 前処理の自動化

- 様々なデータ構造（マルチパート、添付ファイル）のメールに対応した本文抽出

□ 処理の高速化

- 学習プログラムと判定プログラムの分割



7.結論

- 目標 -

本校に蓄積したメールデータを活用し、分析と判定システムの実装を検討



□分析

- 本校に届く迷惑メールは似たようなメールが多く、機械的に送られてきている
- 毎年メールの傾向が変化するため、定期的に学習モデルを変更する必要がある

□判定システム

- 迷惑メール判定を行うシステムの実装
- 件名に迷惑メールとわかるタグをつけた警告メールの送信

8.参考文献

[1]IPA, IPA 情報セキュリティ10大脅威 2021

<https://www.ipa.go.jp/security/vuln/10threats2021.html>

[2]藤森 夏輝, SPAM メールの判別に適した機械学習

<https://www.lab.kochi-tech.ac.jp/yoshilab/thesis/1130378.pdf>

[3]SiTest, 機械学習エンジニア必見！ベイジアンフィルタを実装するための形態素解析ライブラリ3選

<https://sitest.jp/blog/?p=6812>

[4].NET Column, MeCabをPython3で使ってみよう | 形態素解析

<https://www.fenet.jp/dotnet/column/%E8%A8%80%E8%AA%9E%E3%83%BB%E7%92%B0%E5%A2%83/7805/>

[5]100%レンタルサイトを使いこなすサイト, 【Python】メール受信をトリガーとして【さくらVPS】サーバー上の Python プログラムをスタートしたい

<https://go-journey.club/archives/7015>

[6]@charon(TIS株式会社), Janomeを使ってPythonで形態素解析

<https://qiita.com/charon/items/661d9a25b2233a9f8da4>

