SCIS 2011 The 2011 Symposium on Cryptography and Information Security
Kokura, Japan, Jan. 25-28, 2011
The Institute of Electronics, Information and Communication Engineers

ユーザの判断能力に基づいたフィッシングサイト検知技術における一考察

A Consideration for Detection Methods of Phishing Sites based on Users' Ability to Decide

宮本 大輔 * 櫨山 寛章 † 門林 雄基 †
Daisuke Miyamoto Hiroaki Hazeyama Youki Kadobayashi

あらまし 本論文では，ユーザの過去の判断 (Past Trust Decision, PTD) の記録を活用したフィッシングサイト検知についての考察を行う．我々の先行研究では，ユーザがウェブサイトを見て「正規サイトである」「フィッシングサイトである」という判断を行った結果を，既存のヒューリスティクスを用いたフィッシングサイト検知手法に取り入れることを提案している．しかし先行研究では平均的な精度こそ向上していたものの，ユーザによっては彼/彼女らの PTD を用いない場合に精度が向上する場合が確認された．そこで本論文では，PTD を活用すべきユーザ，そうでないユーザについて考察を行う．ユーザを分類する手法として，ユーザの判断能力の構成要素をコンテンツの文言のみで判断を行っていないこと，URL や SSL を確認し怪しさに気付いていること，当該サイトを過去に利用した経験を判断に活用できていることなどあると仮定し，これら要素に基づいたクラスタ分析を行う．また，クラスタ毎に先行研究の提案によるフィッシングサイト検知を行い，活用すべき PTD をもつユーザの傾向について考察を行う．

キーワード フィッシングサイト検知，Trust Decision，機械学習，クラスタ分析

1 はじめに

フィッシング攻撃は，サイバー社会に対する重大な脅威の一つである．この攻撃の特徴は，コンピュータシステムではなく，コンピュータシステムを利用するエンドユーザを標的とする点である．フィッシング攻撃者は，エンドユーザを本物そっくりに作った偽サイトに誘導し，そのウェブサイトに個人情報を入力するよう促す．騙されたエンドユーザがクレジットカードなどの情報を入力してしまうと，その情報が攻撃者に盗み取られる，というのがフィッシング攻撃の手口である．こうした攻撃は被害規模も増えており，市場調査会社の Gartner は調査期間となった 2007 年には約 360 万人がフィッシング攻撃により総額 32 億ドルの損害を被ったと報告している [1]．セキュリティ会社の RSA は，エンドユーザらはフィッシング攻撃を最大の脅威であると捉えていることを報告しており [2]，その対策技術の確立は急務である．

フィッシングサイトの対策技術の 1 つに，ユーザが閲覧しているウェブサイトがフィッシングサイトであるかどうか検知する技術があり，代表的な検知手法としてヒューリスティクス方式が知られている．この方式は，ウェブサイトの URL やドメイン名を分析してフィッシングサイトらしさを計算し，そのスコアによってフィッシングサイトの検知を行う．ヒューリスティクス方式の課題は検知精度であり，新しいヒューリスティクスの開発や，複数のヒューリスティクスの組み合わせ手法などにより，精度の向上を目的とした研究がなされている．

我々の先行研究 [3] では，HumanBoost と名付けた方式を提案した．HumanBoost 方式とは，エンドユーザがウェブサイトに対してこれまで行った，信頼できる，信頼できないといった判断を行った結果 (Past Trust Decision, PTD) を活用し，この PTD の記録を既存のヒューリスティクスと組み合わせるというものである．この研究では，被験者にウェブサイトを表示したブラウザのスクリーンショットを閲覧させ，フィッシングサイトと思うか否かについて質問した．また，ヒューリスティクス方式である CANTINA [4] により各ウェブサイトについてフィッシングサイトか否かを判別させた．この上で，各個人のみで検知した場合の判別誤り率の平均は 20.0%，既存のヒューリスティクスのみで検知した場合は 19.0%，各個人と既存のヒューリスティクスを組み合わせた検知結果を組み合わせた場合は 13.4% であるという結果が観測

* 独立行政法人情報通信研究機構 〒 184-8795 東京都小金井市貫井北町 4-2-1. National Institute of Information and Communications Technology, 4-2-1 Nukuikitamachi Koganei Tokyo,184-8795, Japan. daisu-mi@nict.go.jp
† 奈良先端科学技術大学院大学 〒 630-0101 奈良県生駒市高山町 8916-5. Nara Institutee of Science Technology, 8916-5 Takayama Ikoma Nara, 630-0101, Japan. {hiroa-ha,youki-k}@is.naist.jp

された．しかし，被験者によっては HumanBoost 方式を用いるよりも，既存のヒューリスティクスによって検知させた場合に誤り率が少なくなるケースも確認された．

本論文では，PTD を利用すべきエンドユーザと，そうでないエンドユーザの違いについて考察する．まず，エンドユーザの能力を，ウェブサイトを利用した経験を活用できているかどうか，コンテンツのみによる判断を行っていないかどうか，URL や SSL に基づいた判断ができているかどうか，というような要素によって構成されると仮定する．その上で，各要素について被験者実験に基づいたクラスタ分析を行い，ユーザの分類を行う．その上で，HumanBoost 方式が適応可能なユーザ，そうでないユーザについての比較検討を行う．

実験では 309 人の被験者から解答を集め，能力に応じて 5 個のクラスタに分類した．最も HumanBoost 方式の効果が高かった被験者グループには，利用経験を判断に役立てることができること，ページの内容に頼った判断を行っていないこと，ウェブサイトの URL に基づいた検知を行えること，そしてブラウザの表示するセキュリティ情報を注目できること，といった傾向が観測された．

以下，2 節において関連研究について，3 節において先行研究である HumanBoost 方式について説明する．4 節において被験者実験の概要を述べ，その実施結果に基づいた分析を 5 節に行う．実験の考察を 6 に述べ，まとめと今後の課題について 7 節で述べる．

2 関連研究

2.1 フィッシングサイトの検知方式

フィッシングサイトの検知を行う方式としては，URL フィルタリング方式とヒューリスティクス方式がある．URL フィルタリング方式は，ユーザが閲覧しているウェブサイトの URL を，フィッシングサイトの URL データベースと照合することによって，フィッシングサイトであることを検知する．カーネギーメロン大学において Zhang らが行った 2007 年の調査研究では攻撃の初期段階においてフィッシング検知精度は約 70%であることが示されていた [5]．しかし，同大学で 2009 年に行われた，様々なフィッシングサイトのデータベースを対象とした調査 [6] は，様々なフィッシングサイトのデータベースは，攻撃が行われて間もないフィッシングサイトは，その 20% 未満しかデータベースに登録されていないことを報告した．

ヒューリスティクス方式はウェブサイトの URL やコンテンツなどからフィッシングサイトらしさを計算する方式である．有名なヒューリスティクスの例としては，ドメイン名の取得期間の長さという手法がある．フィッシングサイトは発生してから消滅するまでの期間が短い．従って，ドメイン名が登録されてから現在までの期間が短い場合はフィッシングサイト，そうでない場合は正規サイトというように判別することができる．こうしたヒューリスティクスは必ずしも正確ではないため，複数の異なるヒューリスティクスを組み合わせる必要がある．ヒューリスティクス方式の課題は検知精度にある．前述の Zhang らの調査研究 [5] ではヒューリスティクス方式の SpoofGuard [7] は約 94% のフィッシングサイトを正しく判別できるものの，約 42% の正規サイトを誤ってフィッシングサイトと判別する問題が報告された．

このため，新しいヒューリスティクスの開発により検知精度を高める試みがなされている．Zhang らは，ウェブサイトの文言から重要単語を抽出し，それを検索エンジンに入力し，当ウェブサイトの URL が検索結果の上位に表示されるかどうかで判別するヒューリスティクスを開発し，既存のヒューリスティクスと組み合わせて使うシステム CANTINA を 2007 年に提案した [4]．CANTINA では TF-IDF 値を求める語句はウェブサイトの全単語が対象であったが，2009 年には Xiang [8] らによって，固有表現抽出によってよりウェブサイトの特徴を表す単語を抽出することにより，検知精度を高めるという研究が行われた．

また，ヒューリスティクスの組み合わせ方式を改良することによる検知精度の向上も試みられている．例えば Zhang らの CANTINA では，各ヒューリスティクスに単純な重み付けを行って多数決を行っていた．我々の先行研究 [9] では AdaBoost, SVM, ニューラルネットなど代表的な 9 種類の機械学習手法によるヒューリスティクスの組み合わせを評価した．この研究では，3000 件のウェブサイトを CANTINA で用いられているヒューリスティクスを使って分析し，機械学習による判別及び CANTINA の重み付けによる判別の精度を比較評価した．この研究では精度として判別の誤り率，f_1 値，AUC 値を用いたが，ほとんどの場合において機械学習による手法が CANTINA の重み付け手法を上回った．なお，最も高い精度は AdaBoost [10] の場合において観測された．

2.2 被験者実験の方式

フィッシング攻撃はエンドユーザを狙った攻撃であるため，フィッシング対策技術の有効性を評価するために被験者を募った調査が行われることがある．例えば，フィッシングサイト検知ツールの表示する警告の有効性を調べるため，Wu らは被験者らがツールの検知結果をどう受け取るかという趣旨の実験を行っている [11]．また，エンドユーザを教育することによってフィッシング対策を行うという授業研究 [12,13] では，被験者グループに異なる教材を与えた上で実験を行い，教材の有効性の比較

表 1: PTD データベースの例

URL	Actual Condition	The user's decision	Heuristics #1	···	Heuristics #N
Site 1	phishing	phishing	phishing	···	legitimate
Site 2	phishing	phishing	phishing	···	legitimate
Site 3	phishing	phishing	phishing	···	legitimate
···	···	···	···	···	···
Site M	legitimate	legitimate	legitimate	···	phishing

検討を行っている．

対策技術の評価のために行われる被験者実験とは異なり，エンドユーザがウェブサイトからどのような情報を得ているのかを調査する被験者実験も行われている．代表的な例としては， Dhamija らが 2006 年に行った，22 人の被験者に正規サイト 7 件，フィッシングサイト 13 件を閲覧させた実験が挙げられる [14]．この実験では，フィッシングサイトはインターネットから隔離された環境に再現されており，エンドユーザにブラウザを通してウェブサイトを閲覧させ，エンドユーザの判断の結果及び判断に至るまでの過程を調査している．被験者は男性 10 人，女性 12 人によって構成されており，平均年齢は約 30 歳であった．結果として，23%の被験者らがアドレスバー，SSL などの情報を見逃しており，40%の誤判断を引き起こしていたことが報告された．この他，Fogg らが 2002 年に発表した文献 [15] では，2,684 人を対象としてユーザがウェブサイトの信頼性を何から得ているか調査を行っている．結果として，46.1% のユーザがウェブサイトの見た目から，26.5% が見た目と構造から信頼できるか否かを判断していることが観測されており，著者らはエンドユーザが厳格な判断基準を持っていないことを報告した．

3 HumanBoost

後の議論を正確にするため，先行研究である，HumanBoost [3] 方式について概要を説明する．HumanBoost 方式は，エンドユーザがウェブサイトを信頼できる，信頼できないといった判断を行った結果 (Past Trust Decision, PTD) を，フィッシングサイトの検知に活用するという提案である．エンドユーザはウェブサイトに個人情報を入力する時，つねに何らかの意思決定を行っていると考えられる．言い換えれば，エンドユーザはウェブサイトに対し正規サイトあるいはフィッシングサイトであるという出力を行う装置であるとも考えられる．

仮に，表 1 のように，各エンドユーザに PTD のデータベースが存在していると考える．データベースのスキーマはウェブサイトの URL 及びそのサイトの実際の状況，さらにユーザの意思決定の結果と既存のヒューリスティクスの結果によって構成される．PTD のデータベースを $N+1$ 個の説明変数と 1 個の目的変数を持つバイナリ行列とみなすと，フィッシングサイトの検知は機械学習における分類問題の一種であると捉えられる．

そこで，エンドユーザにウェブサイトを閲覧させ PTD を作成し，PTD を用いた場合，用いない場合の比較検討を行った．機械学習手法には AdaBoost を利用した．理由の 1 つは，単純に 2.1 節に示した通り AdaBoost の性能が高かったためである．他の理由としては AdaBoost は，あるヒューリスティクスが正確に解答できなかったウェブサイトについて，正しく解答できた他のヒューリスティクスに高い重みを割り当てるという理論的背景があるためである．これにより，エンドユーザが間違えやすいフィッシングサイトを正しく判別できるヒューリスティクスに高い重みが割り当てられ，各エンドユーザの能力に応じた検知が行えるのではないかと期待した．なお，複数のエンドユーザが PTD を共有することは，プライシーの問題もあり本論文では考慮しない．

第 1 回の被験者実験は，2007 年 11 月に奈良先端科学技術大学院大学のインターネット工学講座に所属していた 10 名を対象として実験を行った．被験者らは全員 22 歳から 29 歳の男性で，3 人は過去 5 年以内に修士課程を卒業しており，残りは修士課程の学生であった．この実験は 2.2 節で述べた Dhamija の実験を踏襲して行われ，被験者らは Windows XP 上で動作する Internet Explorer (IE) 6 を操作し，正規サイト 6 件，フィッシングサイト 14 件を閲覧した．被験者らの判断の平均の誤り率は 19.0%，既存のヒューリスティクスを AdaBoost で組み合わせた検知の誤り率が 20.0% であったのに対し，HumanBoost 方式の場合は誤り率が 13.4% と改善が見られたことを観測した．なお，特殊な設定として IE 6 を多国語ドメイン名を表示できるようにした事を除いては，OS やブラウザはインストールされたままの標準的な設定を用いた．

また，第 2 回の被験者実験として，2010 年 3 月に北陸先端科学技術大学院大学の篠田研究室に所属していた 11 名を対象として追試を行った．被験者らは全員 23 歳から 30 歳までの男性で，2 人が過去 5 年以内に修士課程を卒業しており，残りは修士課程の学生であった．この実験では被験者らにはブラウザを操作させず，IE 6 のスクリーンショットを紙に印刷したプリントを用いて判断させた．一部の正規サイトは第 1 回目の実験の時からデザインが変更されていた為，フィッシングサイトもそれらに合わせて調整を行った．被験者らの判別の平均誤り率は 40.5%，ヒューリスティクスによる組み合わせは 10.5% であったのに対し，HumanBoost 方式では 9.7% であった．

これらの被験者実験の共通の問題点としては，被験者の偏りが挙げられる．どちらの実験においても，被験者は少数であり，全員男性であり，情報工学分野の修士課程の学生または卒業生であった．また，被験者によって

は PTD を利用せず，既存のヒューリスティクスのみによって判別を行う場合に誤り率が少なくなるケースも確認された．

そこで，本論文では多種多様なエンドユーザらを対象とし，エンドユーザがフィッシングサイトを判断する際の根拠を調査する．そして，そうしたユーザの判断基準に応じてユーザを分類し，PTD を利用すべきユーザの傾向，そうすべきでないユーザの傾向について，被験者実験を通じた調査を行う．

4 被験者実験の概要

第 3 回目の被験者実験として，2010 年 7 月にインターネット調査企業に依頼して 309 人分の解答を採取した．この 309 人のうち，男性は 131 人で，女性は 178 人であった．職業は技術職の会社員が 48 人，事務職の会社員が 58 人，主婦が 61 人，学生が 18 人であった．

設問は，年齢や職業などといったユーザの基本的な属性に加え，ウェブサイトの利用経験の調査，エンドユーザの意思決定の根拠の調査，最後に 3 節で述べた HumanBoost 方式の実験の追試を目的とした調査を目的として設定した．以下，順に説明する．

4.1 ウェブサイトの利用経験についての調査

これまでの被験者実験では，被験者は利用経験の全くないウェブサイトについてもフィッシングサイトか否かの判断を下さねばならなかった．しかし，エンドユーザにとってウェブサイトに事前知識がある場合，ない場合において同様の判断が行えるとは考えがたい．そこで，事前知識として 4.2 節に述べるウェブサイト群について利用経験の有無の調査を行った．

4.2 エンドユーザの意思決定の根拠の調査

2.2 節に述べた通り，過去の被験者実験ではエンドユーザはウェブサイトの内容を意思決定の拠り所としていることが知られている．しかし，フィッシングサイトのページの内容は本物そっくりであり，ページの内容に頼った判断ではフィッシングサイトを正規サイトであると誤って判断する率が高くなると考えられる．

そこで，被験者らにいくつかのウェブサイトのスクリーンショットを閲覧させ，それぞれ正規サイトかフィッシングサイトと思うかを判断させた．また，被験者らにはその判断の根拠が「ページの内容」「ウェブサイトの URL」「ブラウザの表示するセキュリティ情報」「その他」（その他の場合はその事由）の何であったかについて，1 つ以上の選択肢を解答させた．また，実験を行った 2010 年には IE 6 が古く使われなくなっており，OS として Windows Vista，ブラウザとして IE 8 を利用した．IE 8 にはフィッシングサイトを判別し警告を表示する機能があるが，この機能は意図的に外し，エンドユーザにウェブサイトが表示された瞬間のスクリーンショットを閲覧させた．

表 2: 意思決定の根拠の調査に使ったウェブサイト

#	ウェブサイト	真偽	言語	SSL の有無
1	ジャパンネット銀行	真	日本語	SSL (EV SSL)
2	みずほマイレージクラブ	偽	日本語	
3	mixi	偽	日本語	
4	Yahoo! JAPAN	偽	日本語	
5	東京都民銀行	真	日本語	SSL (EV SSL)
6	ガンホー	偽	日本語	
7	Gmail	真	日本語	SSL
8	三菱東京 UFJ 銀行	真	日本語	SSL (EV SSL)
9	三井住友 VISA カード	偽	日本語	SSL
10	twitter	偽	日本語	
11	駅ねっと	偽	日本語	
12	Amazon	偽	日本語	
13	ANA マイレージクラブ	偽	日本語	
14	Ameba	真	日本語	
15	ゆうちょダイレクト	偽	日本語	SSL
16	楽天市場	偽	日本語	
17	スクウェアエニックス	偽	日本語	
18	Goo メール	偽	日本語	
19	ニコニコ動画	偽	日本語	
20	GREE	真	日本語	

表 3: HumanBoost 方式の追試に用いたウェブサイト

#	ウェブサイト	真偽	言語	SSL の有無
1	Live.com	真	英語	
2	東京三菱 UFJ 銀行	偽	日本語	
3	PayPal	偽	英語	
4	Goldman Sachs	真	英語	SSL
5	Natwest Bank	偽	英語	
6	Bank of the West	偽	英語	
7	東京都民銀行	真	日本語	SSL
8	Bank of America	偽	英語	
9	Paypal	偽	英語	
10	Citibank	偽	英語	
11	Amazon	偽	英語	
12	Xanga	真	英語	
13	Morgan Stanley	真	英語	SSL
14	Yahoo	偽	英語	
15	U.S.D of Treasury	偽	英語	
16	三井住友 VISA カード	偽	日本語	
17	eBay	偽	英語	
18	Citibank	偽	英語	
19	Apple	真	英語	SSL
20	PayPal	偽	英語	

この実験に用いたウェブサイト群を表 2 に示す．正規サイトとしてエンドユーザが日常的に利用していそうなウェブサイトを選定した．また，フィッシングサイトにも同様に利用していそうなウェブサイトを模したサイトをインターネットから隔離された環境に作成し，スクリーンショットを取得した．例えばウェブサイト 2 は，みずほマイレージクラブのフィッシングサイトであるが，ドメイン名を正規サイトに似せて作成している．ウェブサイト 4, 17 は実際に報告されたフィッシングサイトであり，ウェブサイト 11, 12 はフィッシングサイトをホスティングすると報告されるなどしたウェブサイトである．この他，SSL を用いない正規サイト及び SSL を用いたフィッシングサイトを用いた．なお，この調査には全ての日本語のウェブサイトを用いた．

4.3 HumanBoost 方式の追試

最後に，第 1 回目及び第 2 回目の実験で用いたウェブサイトのスクリーンショットについても同様に作成した．

表 4: 判断基準と平均誤り率

ページの内容	URL	セキュリティ	誤り率
v			62.1 %
	v		25.5 %
		v	36.9 %
v	v		52.0 %
v		v	60.0 %
	v	v	17.7 %
v	v	v	50.2 %

ウェブサイトの一覧を表 3 に示す．基本的には先行研究 [3] の通りである．なお，追試を目的としているため，スクリーンショットを取得した環境は OS を Windows XP ，ブラウザを IE 6 で統一することにした．このため，一部のウェブサイトを表 2 と重複させ，閲覧環境によって判断の違いがどのように変わるかについても観測することとした．この観測結果については 6 節に述べる．

5 被験者実験の解析

5.1 被験者のクラスタリング

本論文では HumanBoost 方式を利用すべきユーザ，そうでないユーザを分類することを目的として被験者実験を行う．この分類の基準として，フィッシングサイトを正しく判断するための能力をいくつかの構成要素に分解し，それぞれ調査を行うこととした．先述の通り，エンドユーザの意思決定の根拠としては， 4.1 節に挙げたとおり過去のウェブサイトの利用経験， 4.2 節で挙げたとおり，ページの内容，ウェブサイトの URL ，ブラウザの表示するセキュリティ情報などを考える．

そこで，これらの情報がウェブサイトの判断に好影響をもたらすのか，あるいは悪影響をもたらすのかを考える．まず，利用経験がないと答えられたウェブサイトの平均誤り率は 48.6% であったのに対し，利用経験があるサイトでは 42.7% であることが観測された．これから，利用経験の有無は意思決定の結果に何らかの好影響を及ぼしていると推測し得る．

次に，各項目と平均誤り率についての調査を行った．その結果を表 4 に示す．v は該当する選択肢が選択された事を示す．例えば，ページの内容によってのみ判断している被験者の平均誤り率は 62.1% であった．表 4 からは，ページの内容のみによって判断している場合は誤り率が高く，ウェブサイトの URL とブラウザの表示するセキュリティ情報をみて判断している場合に誤り率が少ない．従って，ページの内容に頼って判断することは意思決定の結果に何らかの悪影響を及ぼしており，ウェブサイトの URL とセキュリティ情報をみて判断することは好影響を及ぼしていると考えられる．なお，選択肢として「その他」を定義したが，選択される頻度が少なかったため本論文では分析の対象外とする．

そこで，被験者らがフィッシングサイトを判断するための能力の構成要素は，以下の 5 項目であると仮定する．

要素 1 過去に利用経験のあるウェブサイトについては，この経験を活かした判断を行うこと．

要素 2 ページの内容に基づいた判断を行っていないこと．

要素 3 ウェブサイトの URL に基づいた検知を行っていること．

要素 4 SSL (EV SSL) を利用しているサイトでは，ブラウザの表示するセキュリティ情報に基づいた判断を行っていること．

要素 5 SSL (EV SSL) を利用していないサイトでは，ブラウザの表示するセキュリティ情報に基づいた判断を行っていないこと．

さらに，これらの各構成要素について，$(0 \cdots 1)$ に正規化された範囲内における数値化を試みる．構成要素 1 については，ウェブサイトの利用経験がある場合に，当該サイトについての検知率を算出することとした．例えば 20 サイトのうち被験者が 10 サイトを利用したことがあり，その 10 サイトについて 8 サイトを正しく検知できた場合，この被験者の構成要素 1 についての能力は 0.8 であると定義する．構成要素 2 については，20 サイトについての判断基準の選択肢に「ページの内容」を選ばなかった割合を用い，同様に，構成要素 3 については，選択肢に「URL」を選んだ割合を用いる．構成要素 4 については，SSL を用いている 6 サイト (2 件のフィッシングサイト含む) について「ブラウザの表示するセキュリティ情報」を選んだ割合を用い，構成要素 5 については，SSL を用いてない 14 サイト (2 件の正規サイト含む) について「ブラウザの表示するセキュリティ情報」を選ばなかった場合を用いる．

次にこれらの数値化された 5 個の構成要素に基づいたエンドユーザのクラスタリングを行う．クラスタリング手法の選定として，代表的なクラスタリングのアルゴリズムである EM 法，Fuzzy C means(FCM) 法の比較検討を行った．各アルゴリズムにおいてクラスタ数を設定する手法は多様な方法が考えられるが，我々はエントロピーと純度に基づいてクラスタ数を決定することとした．エントロピーと純度の測定を行うためには，クラスタ数が n の場合，被験者を n 通りに分類する何らかの指標が必要となる．我々は被験者のウェブサイトにおける判断の誤り率を n 個の階級に分けて分類することとした．例えばクラスタ数を 5 とした場合，被験者 i の判断の誤り率 n_i について $n_i < 0.2, 0.2 \leq n_i < 0.4, 0.4 \leq n_i < 0.6, 0.6 \leq n_i < 0.8, 0.8 \leq n_i$ といったような 5 段階に分類する．これにより，クラスタリングの分類結果と誤り

表 5: EM 法, FCM 法によるクラスタ数の比較

クラスタ手法 / クラスタ数	EM		FCM	
	エントロピー	純度	エントロピー	純度
3	0.602	0.734	0.556	0.731
4	0.581	0.663	0.550	0.702
5	0.500	0.699	0.481	0.722
6	0.541	0.589	0.536	0.595
7	0.581	0.505	0.593	0.508
8	0.615	0.508	0.639	0.489
9	0.596	0.518	0.573	0.511
10	0.588	0.472	0.613	0.469

表 6: 被験者グループのクラスタリング

クラスタ ID	被験者数	要素 1	要素 2	要素 3	要素 4	要素 5
1	79	0.193	0.059	0.041	0.050	0.965
2	58	0.814	0.714	0.745	0.196	0.905
3	53	0.791	0.768	0.726	0.723	0.563
4	65	0.511	0.286	0.704	0.130	0.919
5	54	0.460	0.276	0.499	0.183	0.882

率による分類結果を照合し，エントロピーと純度を計算する．

各クラスタリング手法による結果を表 5 に示す．一般に，エントロピーは小さいほど良く，純度は高いほど良いとされる．能力を構成する要素の数を 5 と仮定していることもあり，本論文ではクラスタ数に 5 を選択した．また，EM 法と FCM 法の比較検討の結果，FCM 法によるクラスタリングを採用することとした．

クラスタリングの結果を表 6 に示す．クラスタ ID は便宜上付与した名前であり，それぞれ 5 個のクラスタを意味している．被験者数は各クラスに分類された被験者数である．要素 1〜5 の数字は，被験者の能力の構成要素を 5.1 節で述べた手法により数値化し，クラスタ毎に平均を計算した値である．

5.2 クラスタ毎の HumanBoost 結果

この 5 個の被験者グループについて，各グループの判断の平均誤り率，既存のヒューリスティクスを用いた場合の誤り率，及び HumanBoost 方式を用いた場合の平均誤り率を調査する．まず 4.3 節で述べた通り，20 サイト分の PTD をユーザ毎に作成する．次に，4 分割交差検定法を 10 回繰り返し，誤り率の平均を計算した．

結果を図 1 に示す．白色のグラフが被験者クラスタ毎の判別の誤り率の平均，灰色のグラフがヒューリスティクスによる誤り率の平均，黒色のグラフが HumanBoost 方式を用いた場合の，各被験者クラスタにおける誤り率の平均である．既存のヒューリスティクスによる誤り率は 11.0%，各被験者の誤り率はそれぞれクラスタ 1 が 49.9%，2 が 35.1%，3 が 26.1%，4 が 41.9%，5 が 44.5% であった．はそれぞれ 10.4%，8.8%，8.0%，9.9%，9.9% であった．

被験者の誤り率でいえばクラスタ 3，次いでクラスタ 2,4,5,1 の順番となっている．平均誤り率の低いユーザの PTD であるから，HumanBoost 方式を用いた際に

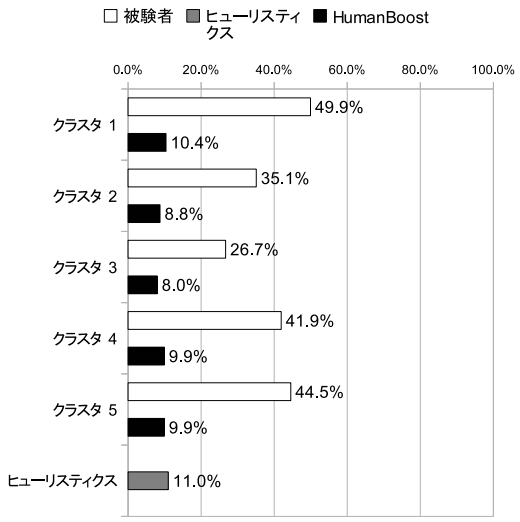

図 1: 各クラスタの誤り率

も平均誤り率が減少しているのは自然であると考えられる．表 6 から読み取れるクラスタ 3 の被験者らは，要素 1,2,3,4 が高い傾向にある．クラスタ 2 の被験者らは要素 1,2,3 が高い傾向にある点が似ているが，要素 4 はクラスタ 3 に比べて低くなっている．クラスタ 1,4,5 はページの内容に頼った判断を行う傾向が強く，HumanBoost 方式による性能の改善もクラスタ 2，3 のユーザに比べるという事象が確認された．

これらの結果から，HumanBoost 方式で活用すべき PTD を持つユーザの傾向として，クラスタ 3 に代表される「事前知識を検知に役立てることができる」かつ「ページの内容に頼った判断を行っていない」「ウェブサイトの URL に基づいた検知を行える」事であり，また「ブラウザの表示するセキュリティ情報を注目できる」といった能力を持つ被験者であると考察できる．

6 考察

4.2 節においてエンドユーザの意思決定の根拠の調査に用いたウェブサイト閲覧環境は OS が Windows Vista，ブラウザが IE 8 であった．また，4.3 では OS が Windows XP，ブラウザが IE 6 と違いがある．このため，表 2，3 に示す通り一部のウェブサイトを同じにし，ブラウザの違いによって影響が出ているのかを観測する．

近代的なブラウザは Extended Validation (EV) SSL 証明書に対応している．EV SSL 証明書は，証明書認証局が対象機関の法的実在性を調査するなど厳格な発行手続きをしているという特徴がある．また，例えば IE 8 の場合は EV SSL 証明書を用いているウェブサイトを閲覧するとアドレスバーが緑色になり，ウェブサイト所有者の名称が表示されるなどされる．Robert らの実験 [16] では，被験者らが EV SSL 証明書に対応した IE 7 を用

表 7: EV SSL 証明書を用いている正規サイトにおける誤り率

ウェブサイト ブラウザ / クラスタ ID	正規サイト (EV SSL) IE 8	正規サイト (EV SSL) IE 6
1	35.4 %	46.8 %
2	60.3 %	79.3 %
3	26.4 %	66.0 %
4	52.3 %	55.4 %
5	51.9 %	51.9 %

いた場合，EV SSL 証明書を使ったウェブサイトの方が通常の SSL を用いたウェブサイトよりもウェブサイトの所有者の情報を発見しやすくなったという結果が示されている．

本論文での実験では，被験者らは，IE 8 と IE 6 という異なる環境において，正規の東京都民銀行のサイトを閲覧している．東京都民銀行のウェブサイトは 2010 年 7 月現在において EV SSL 証明書を用いており， 4.2 節では前述の通りに表示される．しかし， 4.3 節で用いた IE 6 は EV SSL 証明書に対応しておらず，通常の SSL 証明書と同じように鍵アイコンを表示する．なお，東京都民銀行にエンドユーザがログインページは `https://www2.paweb.answer.or.jp/` と東京都民銀行を想起させない URL となっている．さらに，東京都民銀行のウェブサイトの EV SSL 証明書では，ウェブサイトの所有者は NTT DATA CORPORATION と表示されており，被験者らが東京都民銀行のサイトであることを確認することは分かり難かったと考えられる．

しかし，結果として，クラスタ 5 の被験者らを除き，EV SSL 証明書を用いているウェブサイトでは IE 6 よりも IE 8 の方が誤り率が少なくなっている．とりわけクラスタ 3 の被験者らは EV SSL 証明書の有無によって検知率が大幅に改善されている．察するに，クラスタ 3 の被験者らは，ウェブサイトの URL からフィッシングサイトであるかもしれないと考えたのであろう．そして，これらの被験者らは EV SSL 証明書を視認したことによって正規サイトである可能性の方が高くなると意思決定を行ったのではないかと考察し得る．被験者らの意思決定の根拠に用いたウェブサイトと HumanBoost 方式の追試に用いたウェブサイトの閲覧環境を等しくした場合の実験は今後の課題である．

7 まとめ

本論文では，エンドユーザがこれまで行ったウェブサイトについて「正規サイトであり信頼できる」「フィッシングサイトであり信頼できない」といった判断 (Past Trust Decision, PTD) と，既存のヒューリスティクスを機械学習によって組み合わせてフィッシングサイトの判別を行う提案である HumanBoost 方式において，それを活用すべきユーザはどのような傾向にあるのかを調査した．

この調査手法として，被験者にウェブサイトを閲覧させ，その際に何を根拠に判断を行ったかをアンケート形式で実験を行うこととした．まず，被験者らに対し，これから閲覧するウェブサイトについての利用経験を質問した．次に，被験者らに 20 サイトを閲覧させ，フィッシングサイトか否かを判断させた．また，この際にその時に被験者らが活用した情報として,「ページの内容」「ウェブサイトの URL」「ブラウザの表示するセキュリティ情報」及び「その他」を選択させた．次に，先行研究 [3] で用いた 20 サイトを閲覧させ，フィッシングサイトか否かを判断させた．

実験結果として得られた 309 人分の被験者の解答から，被験者が正規サイトとフィッシングサイトを判断する能力は 5 項目の要素によって構成されると仮定した．この上で，被験者らを Fuzzy C Means 法を用い，構成要素に基づいた 5 個のクラスタに分類した．さらに，各クラスタにおける被験者の判断の誤り率，既存のヒューリスティクスを用いた場合の誤り率，被験者の判断と既存のヒューリスティクスを組み合わせた HumanBoost 方式による誤り率の平均値を算出した．既存のヒューリスティクスによる誤り率は 11.0%，5 個のクラスタに分類された被験者グループの誤り率の平均は 49.9%，35.1%，26.1%，41.9%，44.5% であった．HumanBoost 方式による誤り率の平均は，10.4%，8.8%，8.0%，9.9%，9.9% であった．誤り率の改善が高い被験者グループには,「利用経験を判断に役立てることができる」かつ「ページの内容に頼った判断を行っていない」「ウェブサイトの URL に基づいた検知を行える」，また「ブラウザの表示するセキュリティ情報を注目できる」といった傾向が観測された．

ただし，この調査実験ではエンドユーザの判断基準の調査の際に用いたウェブサイトの閲覧環境が Internet Explorer (IE) 8 であったのに対し，追試では IE 6 を用いている．IE 8 と IE 6 では，Extended Validation (EV) SSL 証明書などを利用しているウェブサイトを表示する際に挙動が異なり，エンドユーザが一貫した判断を行えていないという問題が懸念される．また，標本として用いたウェブサイトに対する偏りも考えられる．こうした問題を解決するためには，全く独立した被験者実験の追試を行い，有効性を検証する必要性があると考えられる．

今後の課題としては，被験者実験に用いたブラウザを統一し，また，ウェブサイトに対する偏りを考慮する必要があると考えられる．また，フィッシングの手口が変化した場合においても PTD に基づいた判別が有効か否かを検証し続けることも課題である．

参考文献

[1] Tom McCall. Gartner Survey Shows Phishing Attacks Escalated in 2007; More than $3 Billion Lost to These Attacks. Available at: http://www.gartner.com/it/page.jsp?id=565125, Dec. 2007.

[2] RSA Security, Inc. RSA 2010 Global Online Consumer Security Survey. Available at: http://www.rsa.com/products/consumer/whitepapers/10665_CSV_WP_1209_Global.pdf, Jan. 2010.

[3] Daisuke Miyamoto, Hiroaki Hazeyama, and Youki Kadobayashi. HumanBoost: Utilization of Users' Past Trust Decision for Identifying Fraudulent Websites. *Journal of Intelligent Learning Systems and Applications*, 2(4):190–199, 2010.

[4] Yue Zhang, Jason Hong, and Lorrie Cranor. CANTINA: A Content-Based Approach to Detect Phishing Web Sites. In *Proceedings of the 16th World Wide Web Conference*, May 2007.

[5] Yue Zhang, Serge Egelman, Lorrie Cranor, and Jason Hong. Phinding Phish: Evaluating Anti-Phishing Tools. In *Proceedings of the 14th Annual Network and Distributed System Security Symposium*, Feb. 2007.

[6] Steve Sheng, Brad Wardman, Gary Warner, Lorrie Faith Cranor, Jason Hong, and Chengshan Zhang. An Empirical Analysis of Phishing Blacklists. In *Proceedings of the 6th Conference on Email and Anti-Spam*, Jul. 2009.

[7] Neil Chou, Robert Ledesma, Yuka Teraguchi, Dan Boneh, and John C. Mitchell. Client-side defense against web-based identity theft. In *Proceedings of 11th Annual Network and Distributed System Security Symposium*, Feb. 2004.

[8] Guang Xiang and Jason I. Hong. A Hybrid Phish Detection Approach by Identity Discovery and Keywords Retrieval. In *Proceedings of the 17th World Wide Web Conference*, Apl. 2009.

[9] Daisuke Miyamoto, Hiroaki Hazeyama, and Youki Kadobayashi. An Evaluation of Machine Learning-based Methods for Detection of Phishing Sites. *Australian Journal of Intelligent Information Processing Systems*, 10(2):54–63, 2008.

[10] Yoav Freund and Robert E. Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Science*, 55(1):119–139, 1997.

[11] Min Wu, Rovert C. Miller, and Simson L. Garfinkel. Do Security Toolbars Actually Prevent Phishing Attacks? In *Proceedings of Conference On Human Factors In Computing Systems*, Apr. 2006.

[12] Ponnurangam Kumaraguru, Yong Rhee, Alessandro Acquisti, Lorrie Faith Cranor, Jason I. Hong, and Elizabeth Nunge. Protecting people from phishing: the design and evaluation of an embedded training email system. In *Proceedings of Conference On Human Factors In Computing Systems*, pages 905–914, Apr. 2007.

[13] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason I. Hong, and Elizabeth Nunge. Anti-Phishing Phil: the design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 1st Symposium On Usable Privacy and Security*, Jul. 2007.

[14] Rachna Dhamija, J. Doug Tygar, and Marti A. Hearst. Why Phishing Works. In *Proceedings of Conference On Human Factors In Computing Systems*, Apr. 2006.

[15] Brian Jeffrey Fogg, Leslie Marable, Julianne Stanford, and Ellen R. Tauber. How Do People Evaluate a Web Site's Credibility? Results from a Large Study. Technical report, Stanford, Nov. 2002.

[16] Robert Biddle, P. C. van Oorschot, Andrew S. Patrick, Jennifer Sobey, and Tara Whalen. Browser interfaces and extended validation ssl certificates: an empirical study. In *Proceedings of the 2009 ACM workshop on Cloud computing security*, pages 19–30. ACM, Nov. 2009.