内部犯行を誘発する環境の決定木分析

山田 道洋 †　　新原功一 ‡　　菊池浩明 †

† 明治大学総合数理学部　　‡ 明治大学大学院先端数理科学研究科

1　はじめに

情報セキュリティマネージメントにおいて，内部犯行は最も防止が困難な大きな脅威である．事例も少なく，どのような要因が内部犯行を誘発しているのか観測することは困難である．

そこで，本研究では，内部犯行を誘発または，抑制する要因を明らかにして，マネージメントに活用することを目的とする．共有ゲストアカウントの利用など異なる誘発条件を与えた合計 198 名の被験者に単純なタスクを行わせ，内部犯行の有無を調査した．この実験結果について，（1）内部犯行を行う条件を表す決定木，（2）不正を表す連関規則の抽出を行う．

2　提案方式

2.1　目的

新原らは，第三者による監視が低い場合に不正事象が発生する確率が高くなるという実験結果を報告している [1]．Hausawi は，専門家にヒアリングを行い，最も否定的な行為は認証情報（パスワード等）の共有であることを報告している [2].

本研究は，共有アカウントの利用環境にて内部犯行を誘発，抑制する要因を特定することを目的とする．

2.2　実験概要

2016 年 10 月 31 日〜11 月 2 日にクラウドソーシングサイトで募集した 198 名を被験者とし，被験者を共有ゲストアカウント (ID: guest) と個別アカウントの 2 つのグループに自動的に振り分けて，筆者らが作成した WEB サイトにて検索エンジンを評価する単純作業を行ってもらい，不正の有無を観測する．

2.3　共有アカウントの利用

共有アカウントには個別アカウントよりも管理コストが低く済むという利点がある．これは利用者が増えた場合などでも新たに ID やパスワードを発行する必要がないからである．しかし，共有アカウントは個別アカウントと比べて内部犯行を誘発しやすいと言われている．これは，共有アカウントの場合作業者の特定が困難で，不正がばれないという心情にさせてしまうからである．

2.4　作業の流れ

共有アカウントグループの被験者は ID などの入力をすることなくゲストアカウントとして作業手順の確認へと進む．その後，提示されたワードリストの内 50 語以上のワードを検索して結果を報告する．

2.5　不正事象

被験者が検索サイトにて検索したワードが，50 語未満であった場合に不正事象とみなし，「作業の途中放棄」と定義する．検索作業は 50 語以上の検索を行わなくても作業の完了報告ができる様にし，不正行為が一定数発生することを期待する．

3　実験結果

3.1　実験結果

被験者の各属性とアカウントのグループ毎の不正者数，総人数を表 2 に示す．共有アカウント利用グループの方に，より多くの不正者が発生した．

3.2　決定木

決定木は，ターゲットである属性を決定する論理条件を明らかにする機械学習であり，根に近い属性が最も大きな条件となる属性である．「途中放棄」をしたか否かをターゲット属性として，R のパッケージ “rpart” により学習した決定木を図 1 に示す．ここで，「Age>=55」等の分岐の条件を各節点の上に示し，左側の枝が条件にあてはまる．「不正者数/正規者数」を各節点の下に示す．“malicious” や “ok” は人数の多い方を示す．例えば，図

Decision tree analysis on environmental factors of insider threat.

†Michihiro Yamada　‡Koichi Niihara　†Hiroaki Kikuchi

†Meiji University Undergraduate School of Interdisciplinary Mathematical Sciences

‡Meiji University Graduate School of Advanced Mathematical Sciences

表 1　抽出された連関規則 (一部)

No.	lhs(条件部)	rhs(結論部)	support	confidence	lhs.support	lift
1	{group=個別,job=自営業 } ⇒	{Judge=ok}	0.1313131	0.8965517	0.1464646	1.089063
2	{group=個別,Age=40's} ⇒	{Judge=ok}	0.1717172	0.8947368	0.1919192	1.086858
3	{group=個別,Age=30's} ⇒	{Judge=ok}	0.1868687	0.902439	0.2070707	1.096214
4	{group=個別,Sex=Male,job=自営業 } ⇒	{Judge=ok}	0.1111111	0.9166667	0.1212121	1.113497
5	{group=共有 } ⇒	{Judge=malicious}	0.1010101	0.2040816	0.4949495	1.154519

表 2　各属性とアカウントのグループ毎の不正者数と総数

	共有アカウント		個別アカウント		合計	
グループ	不正者	*N*	不正者	*N*	不正者	*N*
男性	13	51	11	58	24	109
女性	7	47	4	42	11	89
19 歳以下	1	1	0	0	1	1
20 歳～29 歳	2	15	2	8	4	23
30 歳～39 歳	9	35	4	41	13	76
40 歳～49 歳	2	30	4	38	6	68
50 歳～59 歳	2	12	2	10	4	22
60 歳～	4	5	3	3	7	8
会社員	5	22	5	26	10	48
公務員	1	1	0	0	1	1
自営業	7	28	3	29	10	57
パート，アルバイト	1	9	0	10	1	19
専業主婦，専業主夫	2	19	2	18	4	37
学生	1	1	1	1	2	2
無職	1	9	3	12	4	21
その他	2	9	1	4	3	13
合計	20	98	15	100	35	198

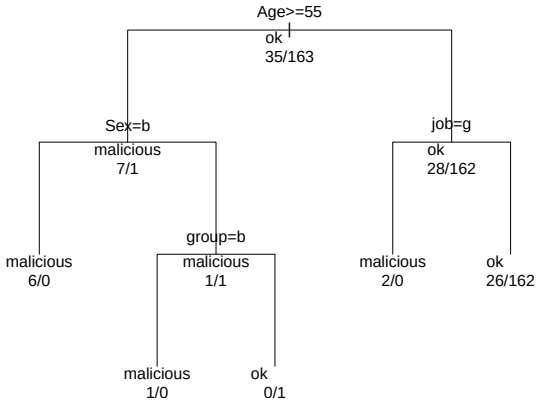

図 1　「途中放棄」の決定木

1 の木では年齢が 55 歳以上かどうかが不正を犯すかどうかを決める最も大きな条件であり，55 歳以上の内 7 名の不正者と 1 名の正常者がいる．

3.3　連関規則

実験結果から属性の組み合わせによる不正への影響を明らかにする為に，R のパッケージ “arules” により連関規則の抽出を行った．抽出した連関規則の一部を表 1 に示す．

support（支持度）は同時確率 $p(lhs, rhs)$，すなわち，実験全体で条件部 lhs と結論部 rhs が同時に起こる確率である．confidence（確信度）は lhs で条件付けられた rhs の条件付き確率 $p(rhs|lhs)$，すなわち，lhs の属性の組み合わせを持つ被験者の中で rhs が発生する確率である．例えば，No.1 の規則は，「個別グループ，かつ，職業が自営業の被験者は 89% の確率で正規者」であることを意味している．

No.5 の規則は，「共有アカウントグループの場合に不正を犯す」を表す．また，個別アカウント単体から成る規則は抽出されなかったが，No. 1 ～ 4 のように，個別アカウントを利用していた特定の職業・年齢の属性を持つ被験者は不正を犯しにくいという規則が代わりに抽出された．

4　まとめ

決定木により，不正を誘発する要因として年齢が大きいことが示された．また，連関規則により，共有アカウントならば，20% の確率（confidence）で不正を犯す規則が抽出された．本稿では省略したがロジスティック回帰分析により個別アカウントは，共有アカウントよりも不正発生を 68% に抑制することが示されている．

参考文献

[1] 新原功一，菊池浩明 : e ラーニングをモデルとした内部犯行の予測因子の識別，Computer Security Symposium 2015, 情報処理学会，pp. 747-754, 2015.

[2] Hausawi, et. al，Current Trend of End-Users' Behaviors Towards Security Mechanisms，HAI，HCI 2016，pp．140-151，2016.