

K406 菊池研・斉藤研合同発表会 2017年2月4日

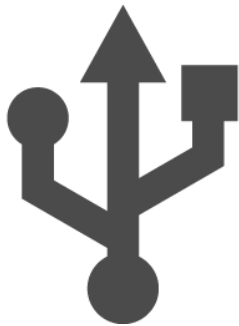
## 内部犯行を誘発する環境の機械学習による分析

明治大学総合数理学部

菊池研4年 山田 道洋

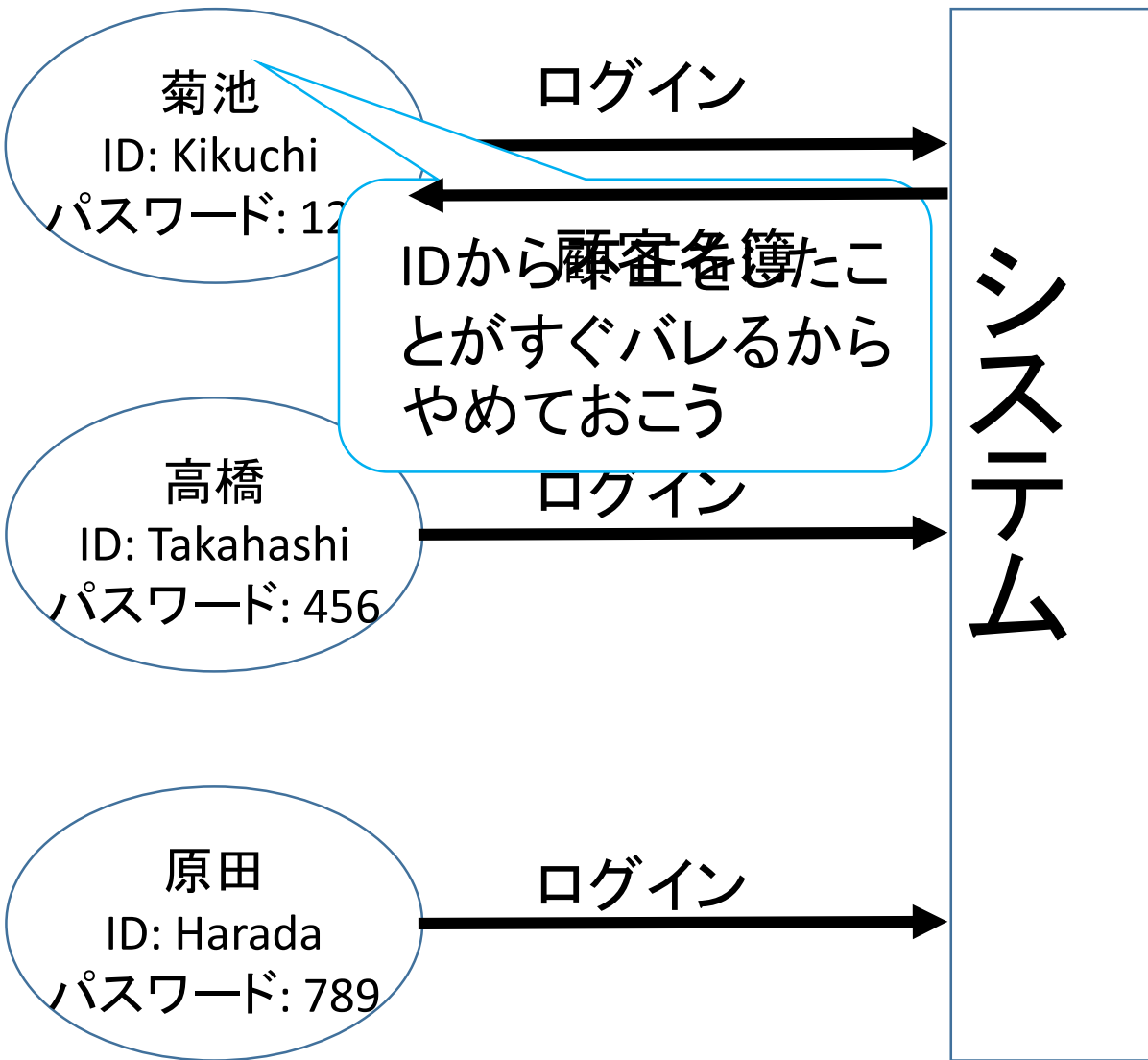
# 研究背景

- 情報のデジタル化により、機密・個人情報の持ち出しが容易になっている
- 2015年にはベネッセ社において、内部犯行による顧客情報の流出事件が発生している
- 内部犯行を誘発する要因には様々なものが挙げられているが、それぞれの要因にどれほどの影響があるかはわかっていない

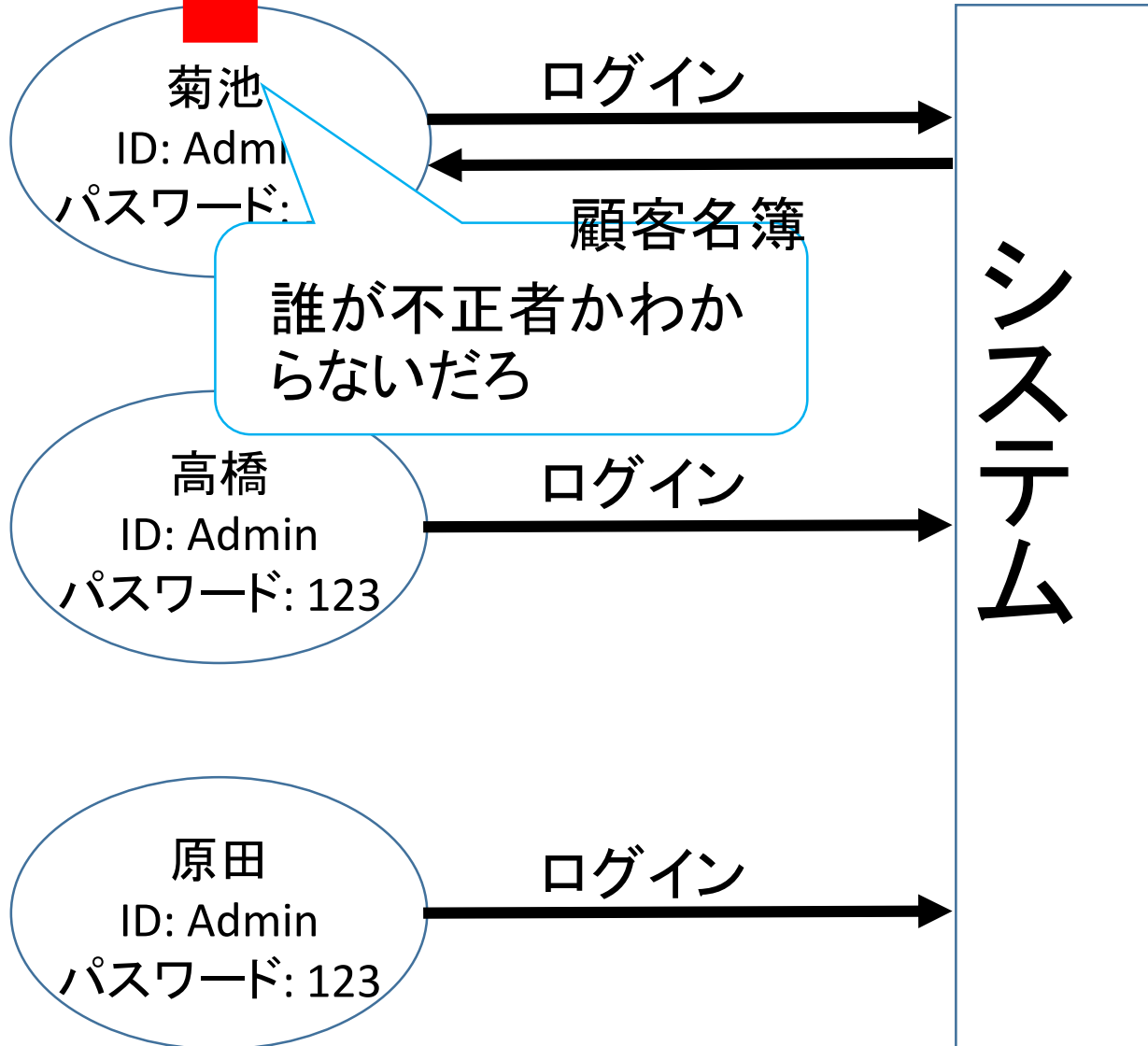


# 内部犯行誘発要因例

個別アカウント



共有アカウント



# 研究目的

- 内部犯行を誘発、抑制する要因を特定すること

# 研究課題

- 実環境では内部犯行の発生数が少ない
- 実環境にて観測された犯行の過程を公開することは企業のセキュリティポリシーに抵触する可能性がある

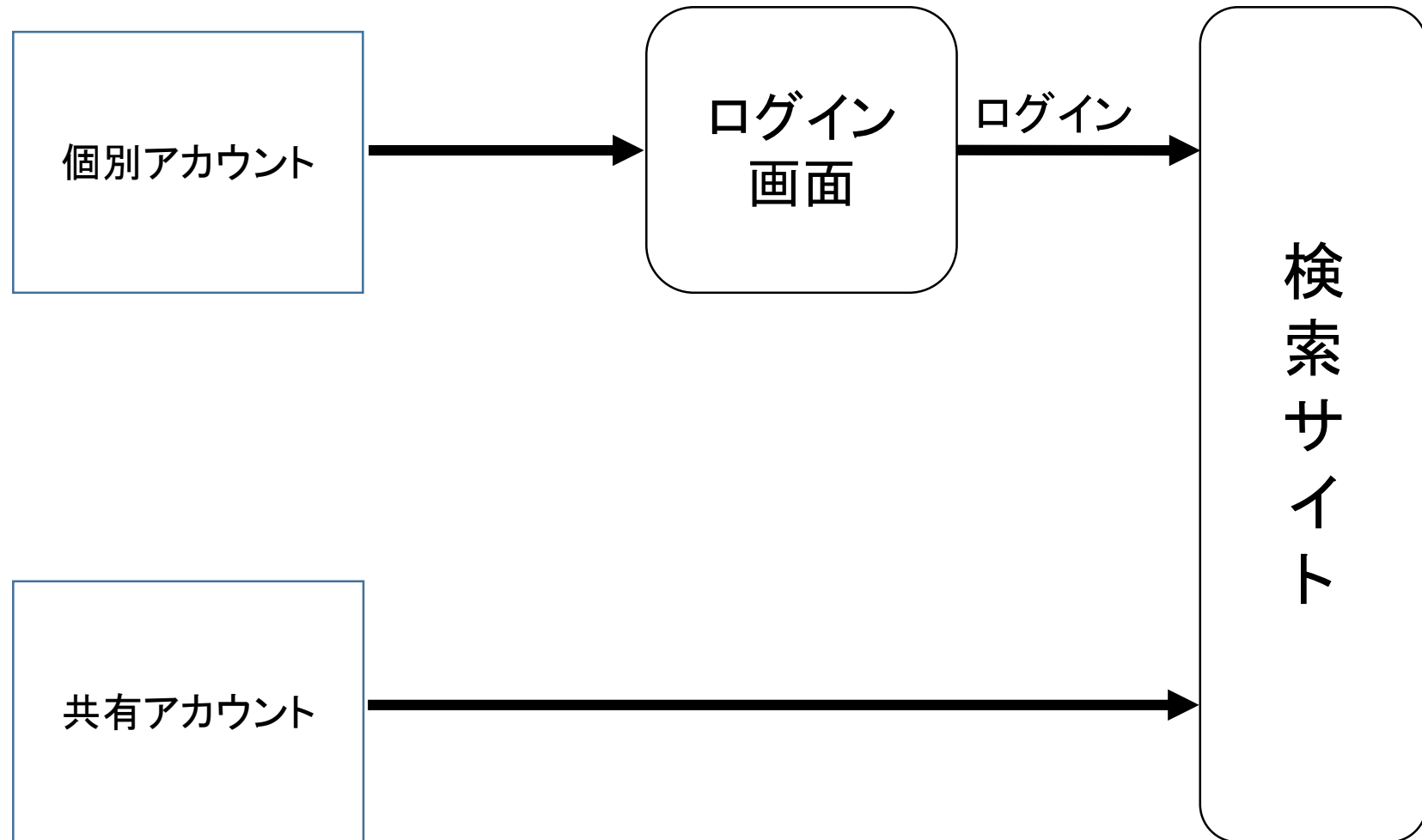
# 解決策

- 自作の環境にて実験を行うことで、作業の過程を記録する
- 不正行為を誘発させる要因を被験者に付与することで不正を一定数発生させる

# 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
- 198名の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つのグループに振り分ける
- 自作の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

# 共有アカウントと個別アカウントの違い



# 実験サイト実行例

## 検索キーワードリスト

後席	急使
手答	手間
揚句	支出
新橋	早目
椿事	此奴

ユーザID: aさん



約 247,000 件(0.84 秒)

Ads by Google

[Amazon後席](#)  
[www.amazon.co.jp/car](http://www.amazon.co.jp/car) ▼  
お急ぎ便なら当日、翌日お届け可能。プライムなら、対象商品は配送料無料  
スマホアプリあり・一億種の品揃え・PC、携帯でいつでも買い物  
サービス: Prime なら動画見放題、100万曲以上聴き放題、通常配送無料

カー用品      バーゲン  
レーダー探知機      ヘルメット

車の「後席」はなんと読みますか？ - 「こうせき」が正しい、自動車技術会が ...  
[detail.chiebukuro.yahoo.co.jp/qa/question\\_detail/q1425683889](http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q1425683889)

Yahoo!知恵袋  
○「こうせき」が正しい、自動車技術会が発行する『日英中自動車用語辞典』(ISBN978-4-915219-70-2)によれば、「こうせき」と読むのが正しいようです。辞典には、「後席」という言葉は掲載され ...

- 50語以上のワードの検索を依頼
- 50語未満の検索は不正行為
- サイト上には検索回数などは表示しない

## 実験結果(サマリー一部)

	共有アカウント	個別アカウント	合計
男性	13	11	24
女性	7	4	11
19歳以下	1	0	1
20歳～29歳	2	2	4
30歳～39歳	9	4	13
40歳～49歳	2	4	6
50歳～59歳	2	2	4
60歳～	4	3	7
合計	20	15	35

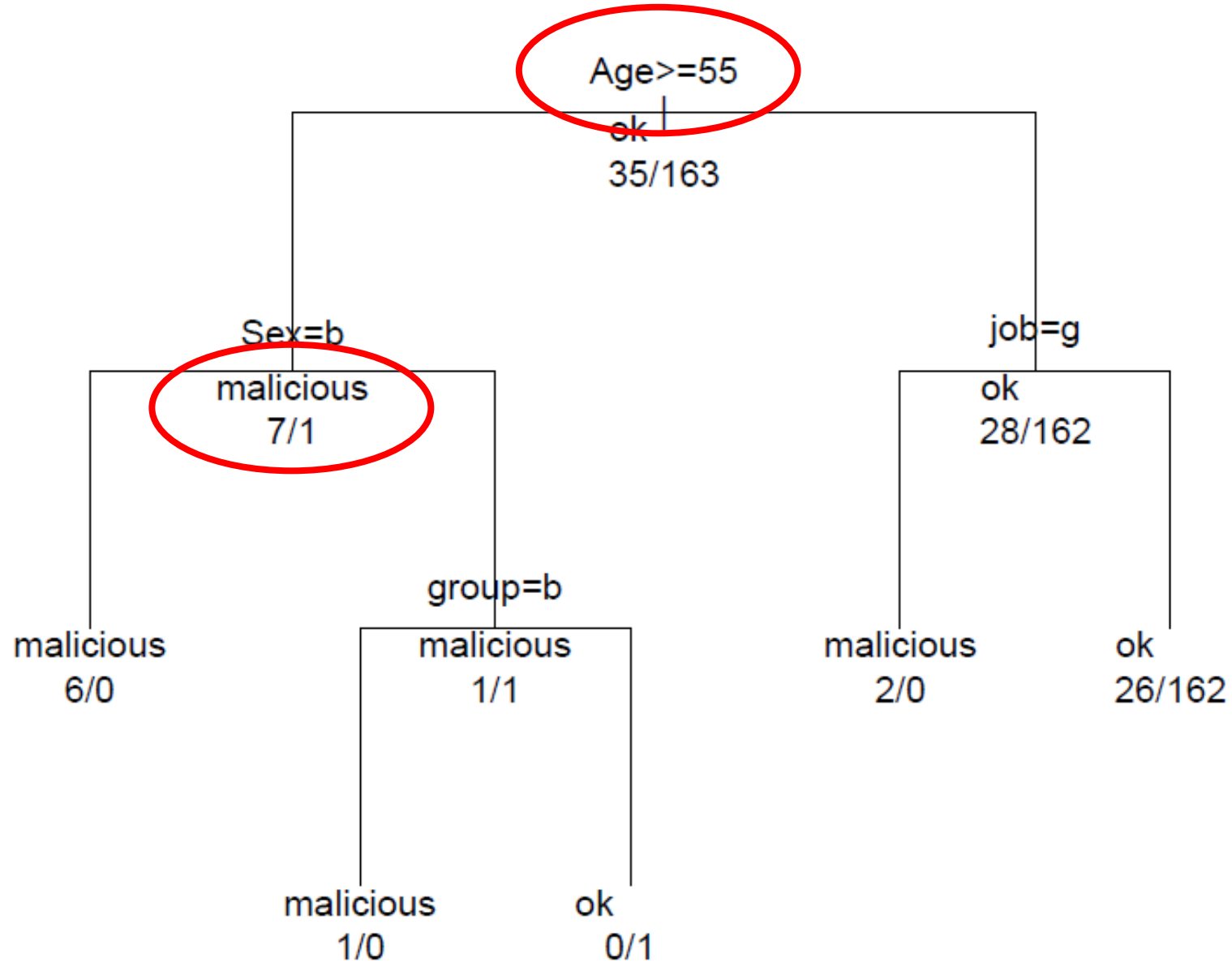
	不正者	正規者
共有	20	78
個別	15	85



# 分析手法

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
3. 各属性の不正に対するオッズ比を求めるロジスティック回帰の分析

# (1)不正を犯したかどうかの決定木



## (2) 不正を犯したかどうかの連関規則(一部)

No	lhs(条件部)	rhs(結論部)	support	confidence	lift
1	{group=個別,job=自営業}	{Judge=ok}	0.131	0.896	1.089
2	{group=個別,Age=40's}	{Judge=ok}	0.171	0.894	1.086
3	{group=個別,Age=30's}	{Judge=ok}	0.186	0.902	1.096
4	{group=個別,Sex=男,job=自営業}	{Judge=ok}	0.111	0.916	1.113
5	{group=共有}	{Judge=malicious}	0.101	0.204	1.154

support(支持度)  $p(\text{lhs}, \text{rhs})$  : rhsとlhsの同時確率

confidence(確信度)  $p(\text{rhs} | \text{lhs})$  : lhsで条件づけられたrhsの条件付き確率

lift(改善率)  $p(\text{rhs} | \text{lhs}) / p(\text{rhs})$  : ターゲットとするrhsが全体で発生する確率に対するlhsの条件付き確率確率がどれだけ向上するかの値

### (3) ロジスティック回帰分析

	Estimate	Pr(> t )	Odds
(Intercept)	-0.107	0.384	2.41E-02
group個別	-0.054	0.306	6.78E-01
Sex男性	0.048	0.466	1.41E+00
Age	0.006	0.024	1.05E+00
job自営業	0.031	0.736	1.38E+00
job会社員	0.097	0.298	2.18E+00
jobその他	0.087	0.476	1.86E+00
jobパート	-0.06	0.567	4.41E-01
job公務員	0.668	0.082	2.90E+07
job学生	1.012	0.000	3.37E+08
job無職	0.064	0.559	1.74E+00

# まとめ

- 内部犯行の発生に影響を及ぼす要因の分析を行った
- 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
- 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
- ロジスティック回帰分析から個別アカウントの利用により共有アカウント利用時に比べて約33%不正を抑制できるという結果が示された