# 類似度に基づいた評価データの選別による マルウェア検知精度の向上

FFRI, Inc.
株式会社 FFRI
http://www.ffri.jp

# はじめに

- このスライドは CSS/MWS 2013 の発表資料です
  - http://www.iwsec.org/css/2013/english/index.html
- 詳細なデータについては、元の論文を参照ください
  - http://www.ffri.jp/assets/files/research/research_papers/MWS2013_paper.pdf
- 質問、コメント等は下記までお願いします
  - research-feedback@ffri.jp

# アジェンダ

- 背景
- 課題
- 本研究の目的
- 実験1
- 実験2
- 実験3
- 考察
- まとめ

# 背景 – マルウェア及び対策技術の現状

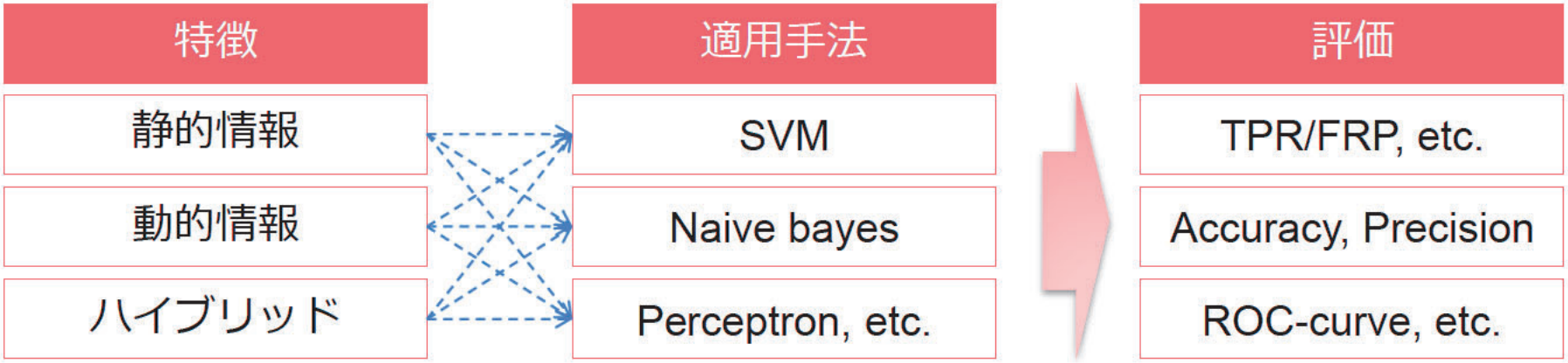
マルウェア生成ツール
難読化ツール
→ マルウェアの急増
標的型攻撃/未知検体
→ パターン方式の限界

新たな検知方式

ビッグデータ →
- ヒューリスティック
- 機械学習
- クラウド レピュテーション

# 背景 – 関連研究の概観

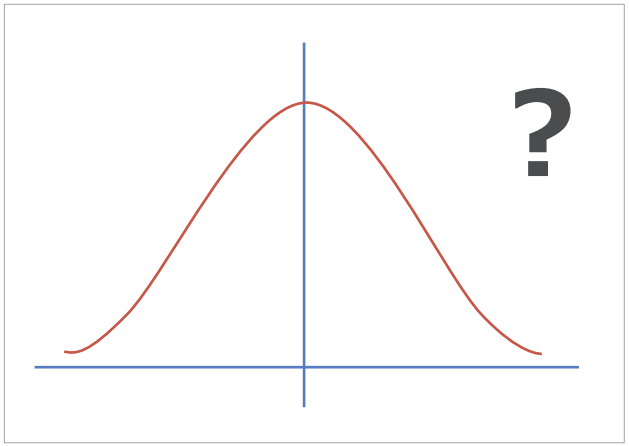
- 下記要素の組み合わせ、工夫が中心
  - 採用する特徴、特徴の加工方法、各種パラメーター設定等
- TPR90%超, FPR1%未満等、比較的良好な結果も

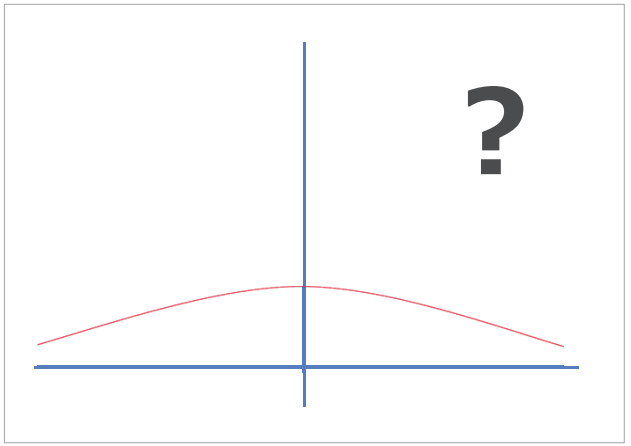


# 課題

- 機械学習一般において
  - 学習データ/評価データの傾向が著しく異なる場合、分類精度が著しく低下する**（結果はデータ次第）**

- マルウェア/正常系ソフトウェアにおいてはどうか?
  - 類似性の分布が広い
    → 学習/評価データで傾向が異なる可能性大



# 本研究の目的

機械学習によるマルウェア検知の有効性を検討する

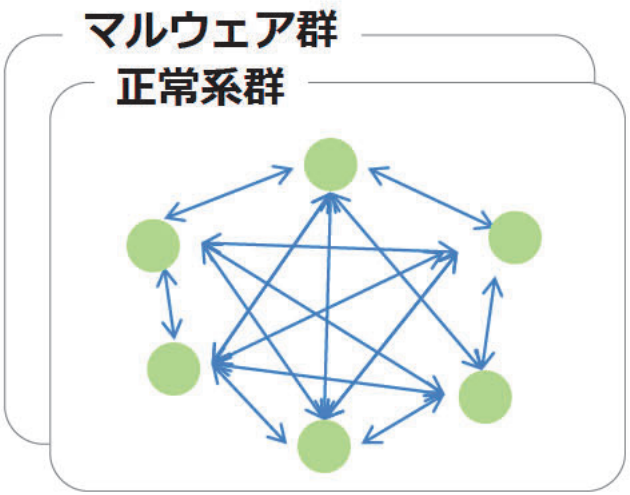
①マルウェア群/正常系ソフトウェア群の分布(類似度)傾向について調査する（実験*1*)

②分布の違いによる分類精度への影響を調査する(実験*2*)

③上記結果に基づき、学習データから類似度が低いデータを除外した場合の分類精度の変化について調査、考察する(実験３)

# 実験1(1/3) – 実験方法

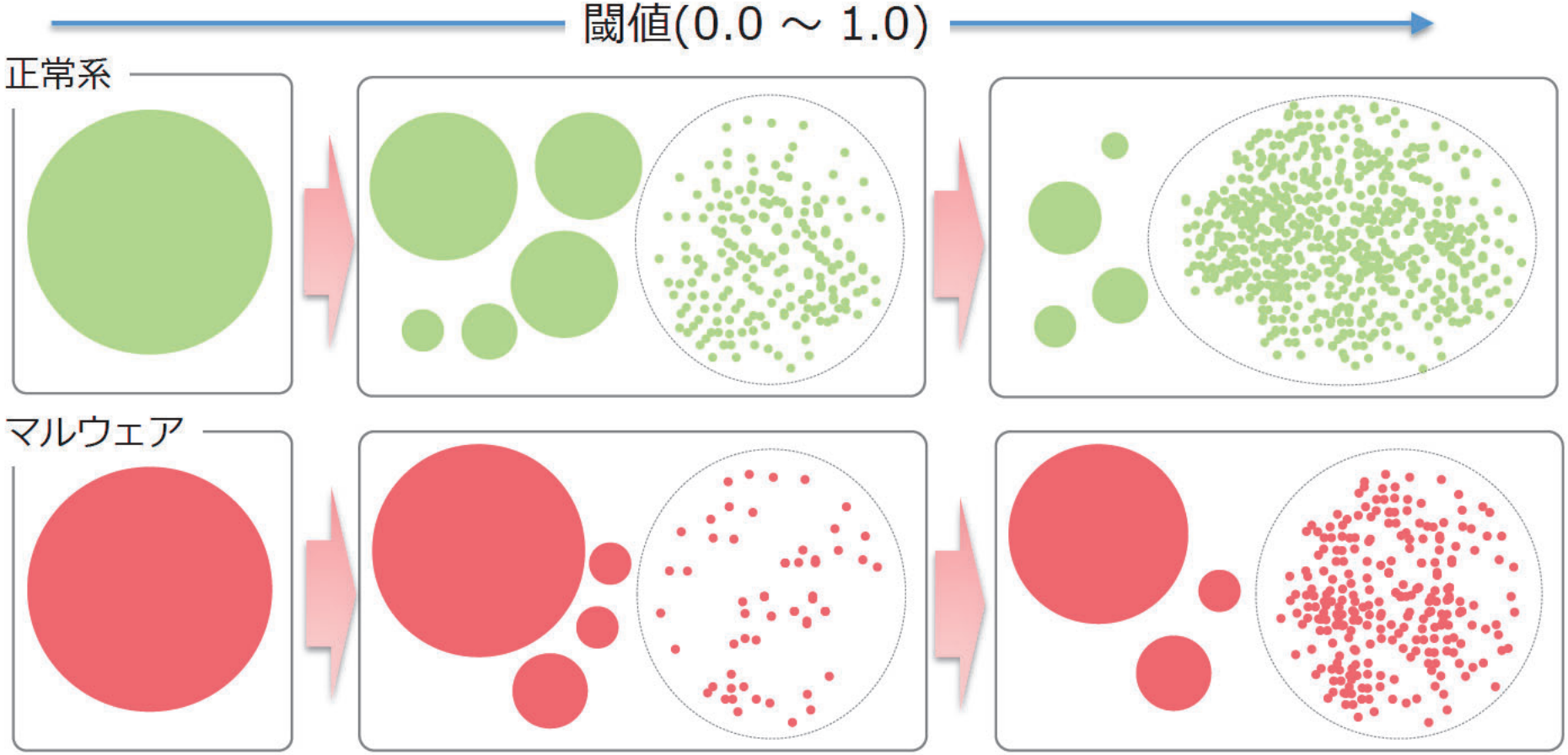
- FFRI Dataset 2013 / 独自に用意した正常系ソフトウェアを利用
- マルウェア/ 正常系同士の相互の類似度を算出 (Jubatus, MinHash)
- 時系列でのAPIコールの4-gramを特徴として利用
  - 例: *NtCreateFile_NtWriteFile_NtWriteFile_NtClose: n回*
    *NtSetInformationFile_NtClose_NtClose_NtOpenMutext: m回, etc.*




| | A | B | C | ... |
|---|---|---|---|---|
| A | — | 0.8 | 0.52 | ... |
| B | — | — | 1.0 | ... |
| C | — | — | — | ... |
| ... | — | — | — | — |

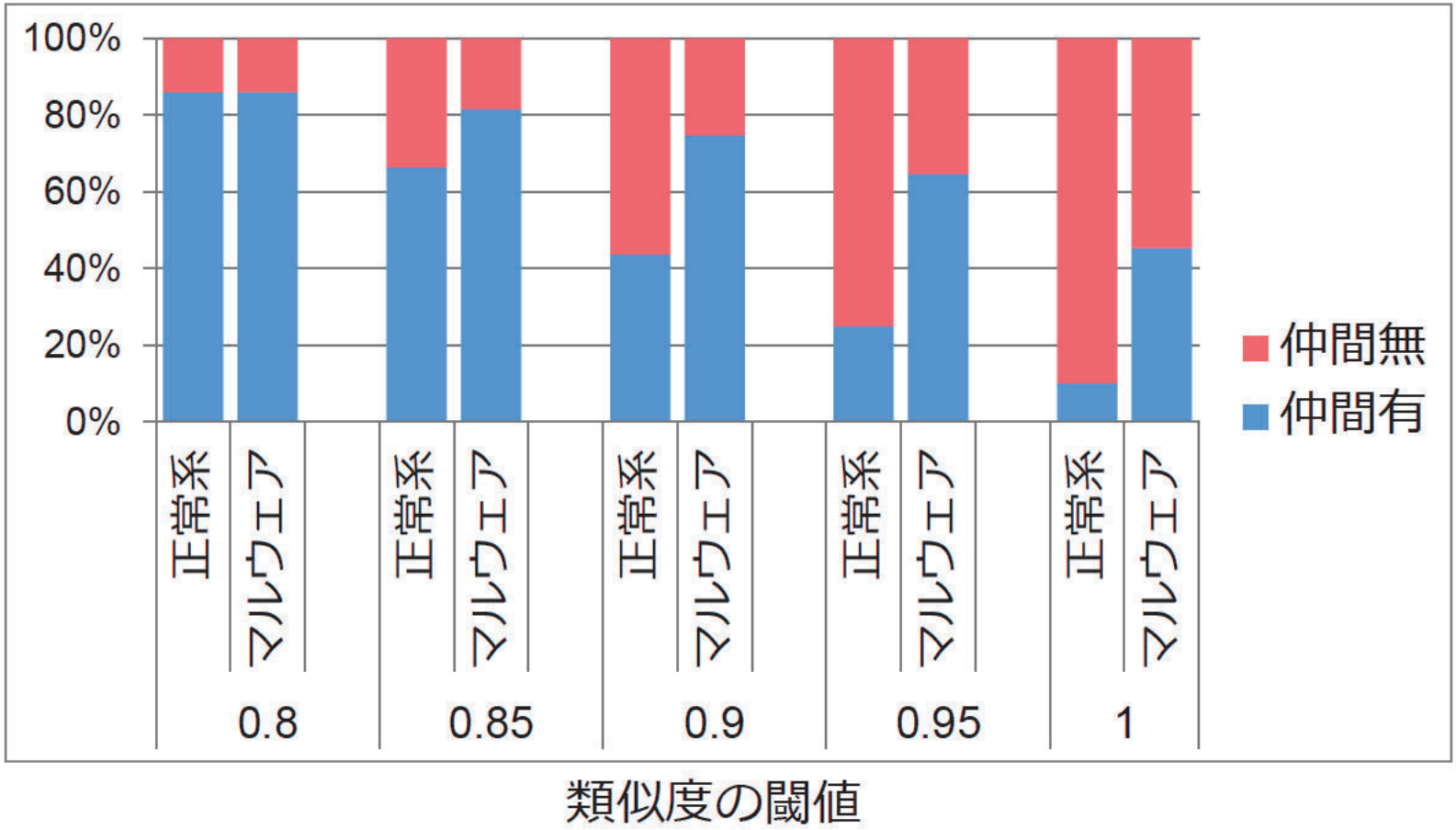
# 実験1(2/3) – 実験方法

- 類似度に閾値を設けてデータをグループ化



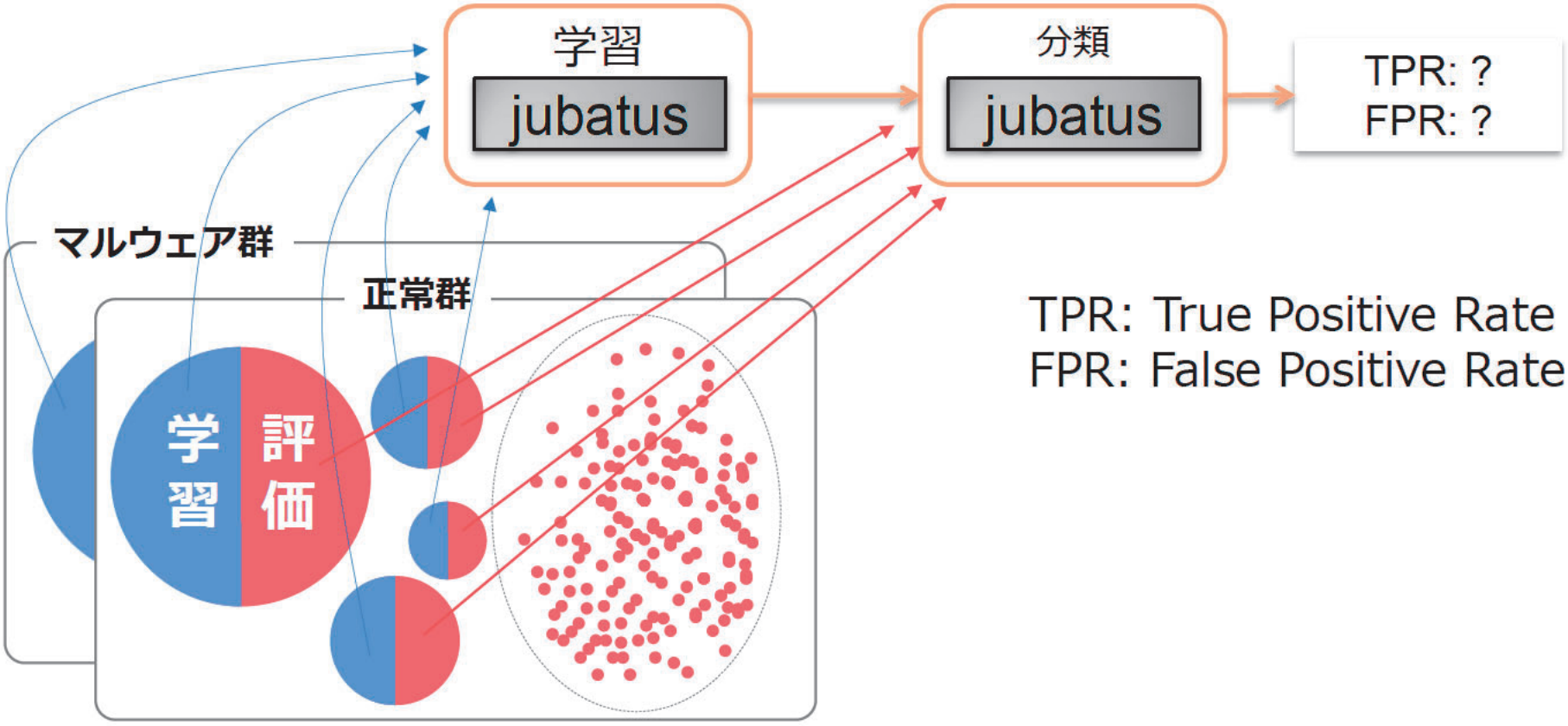
# 実験1(3/3) – 実験結果

## 正常系は、マルウェアに対して似たものを見つけ難い



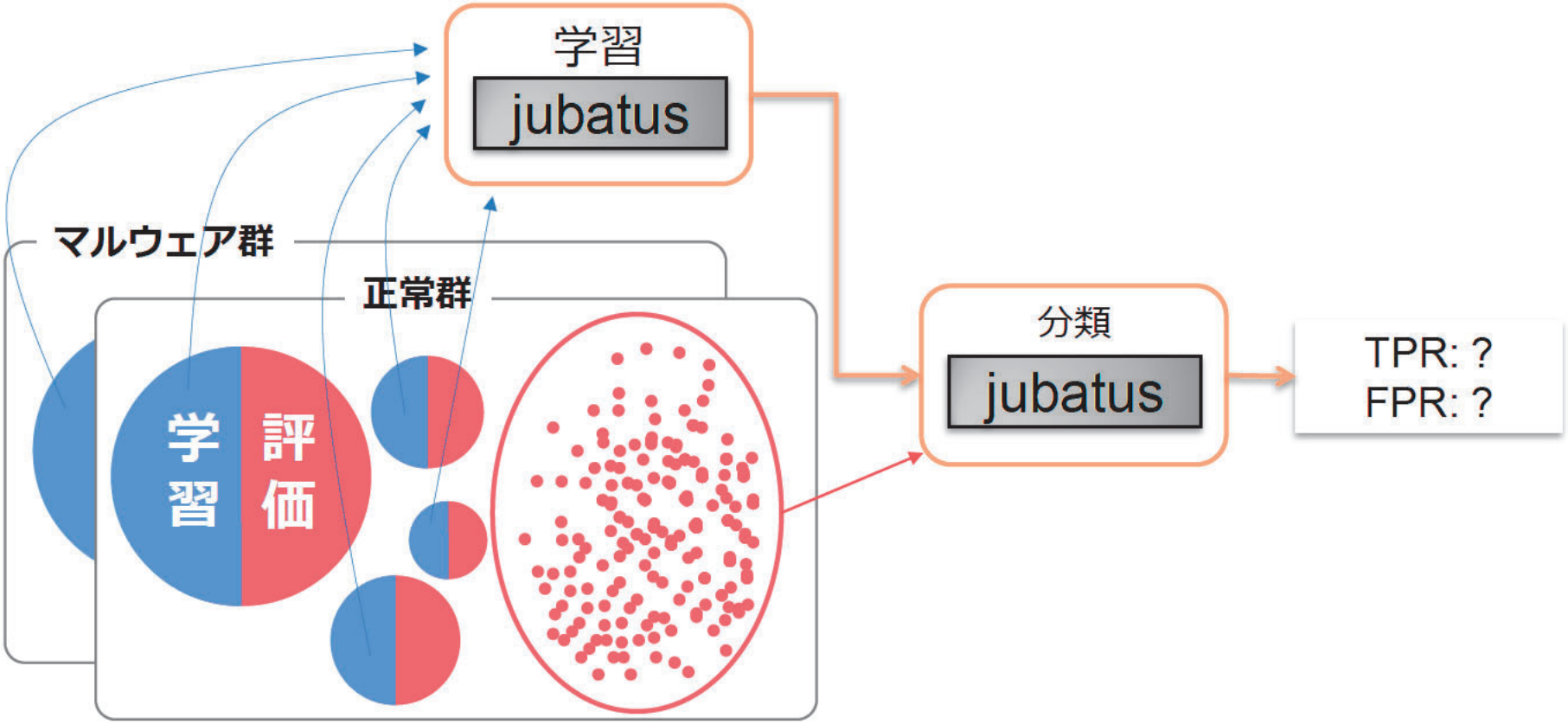
# 実験2(1/3) – 実験方法

- 学習-評価データの傾向差異は、分類精度にどの程度の影響を与えるか?
- 閾値0.9の状態を2分割し、学習・分類を実施(Jubatus, AROW)



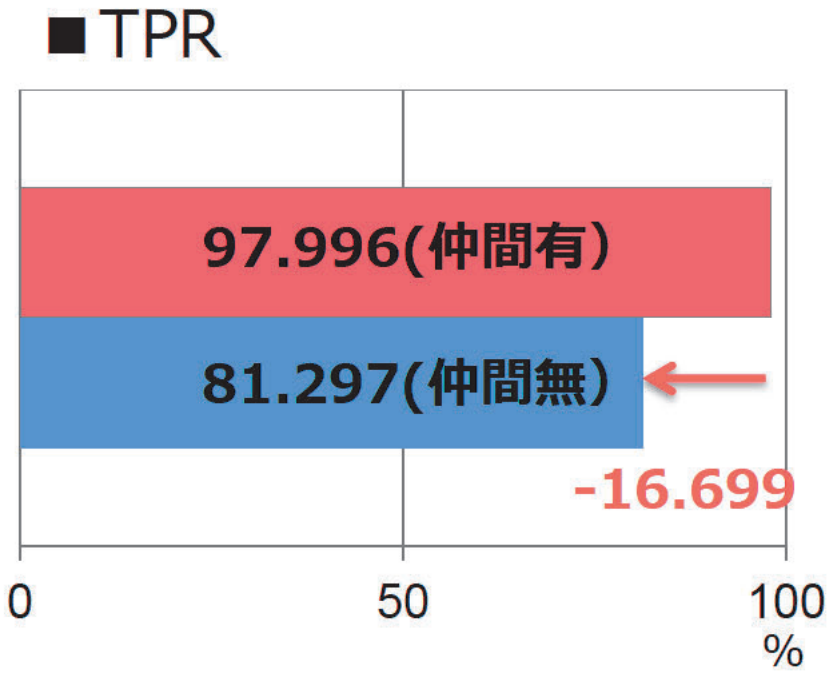
# 実験2(2/3) – 実験方法

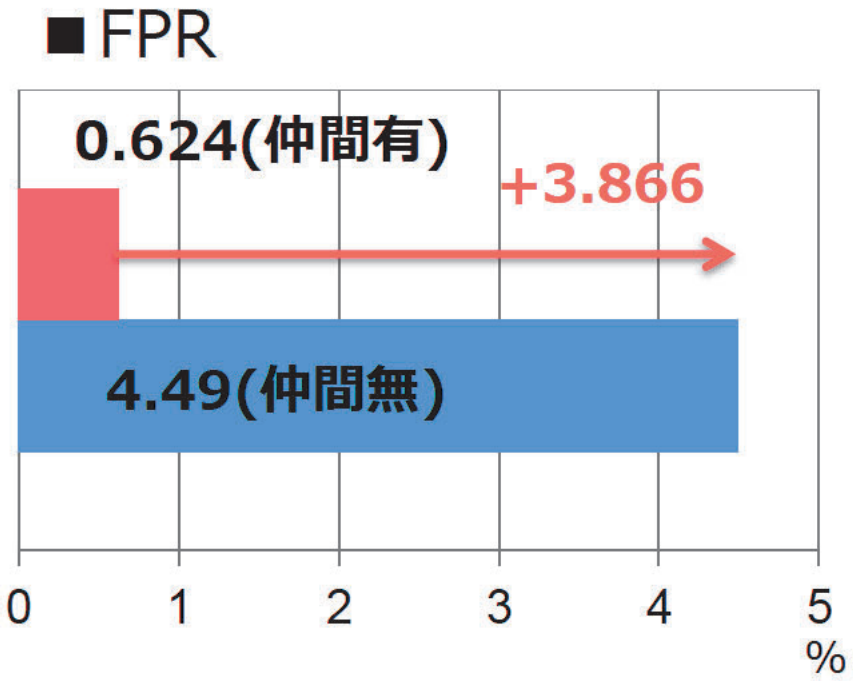
- 学習-評価データの傾向差異は、分類精度にどの程度の影響を与えるか?
- 閾値0.9の状態を2分割し、学習・分類を実施(Jubatus, AROW)



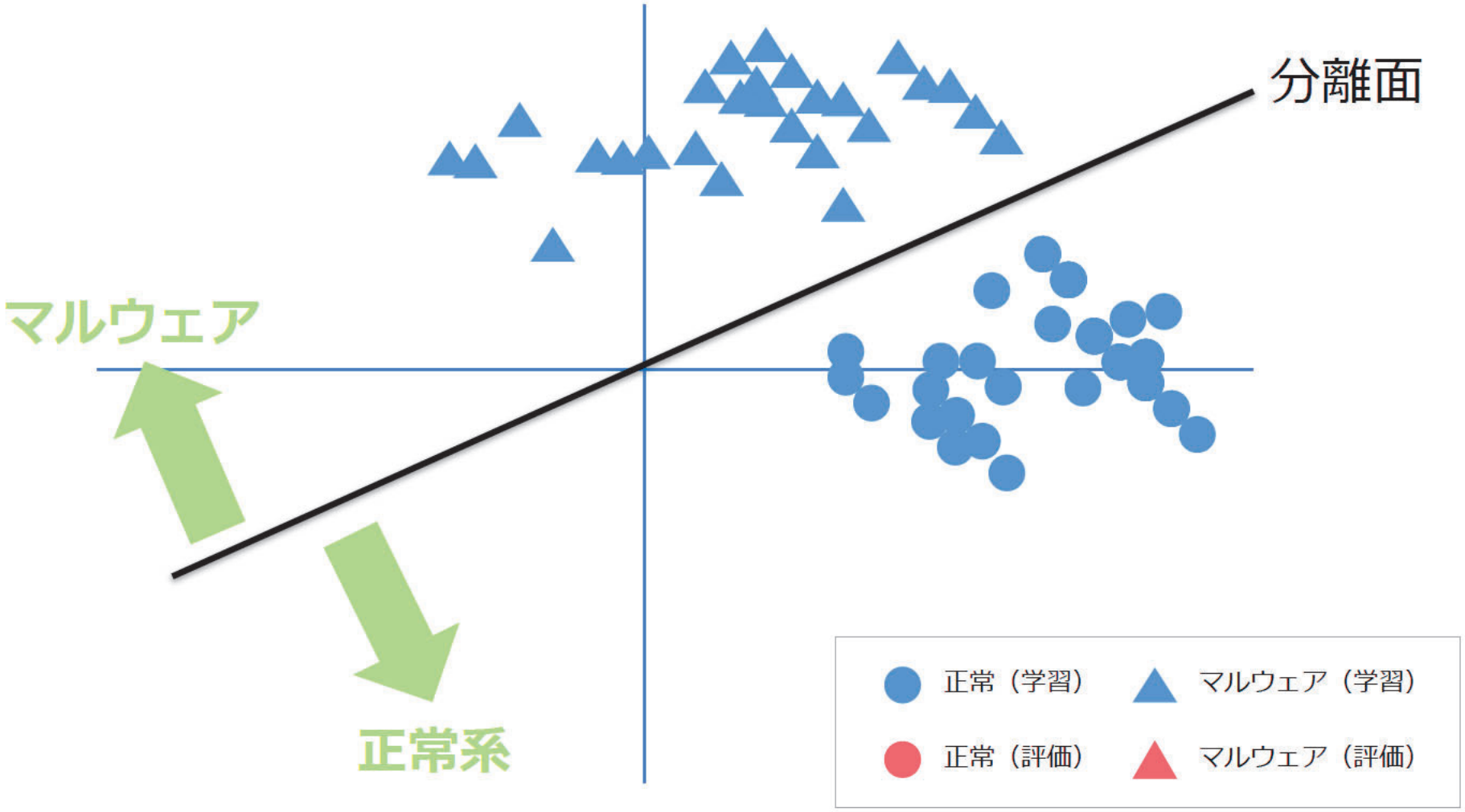
# 実験2(3/3) – 実験結果

## 学習データ/評価データの傾向が異なると分類精度が低下する

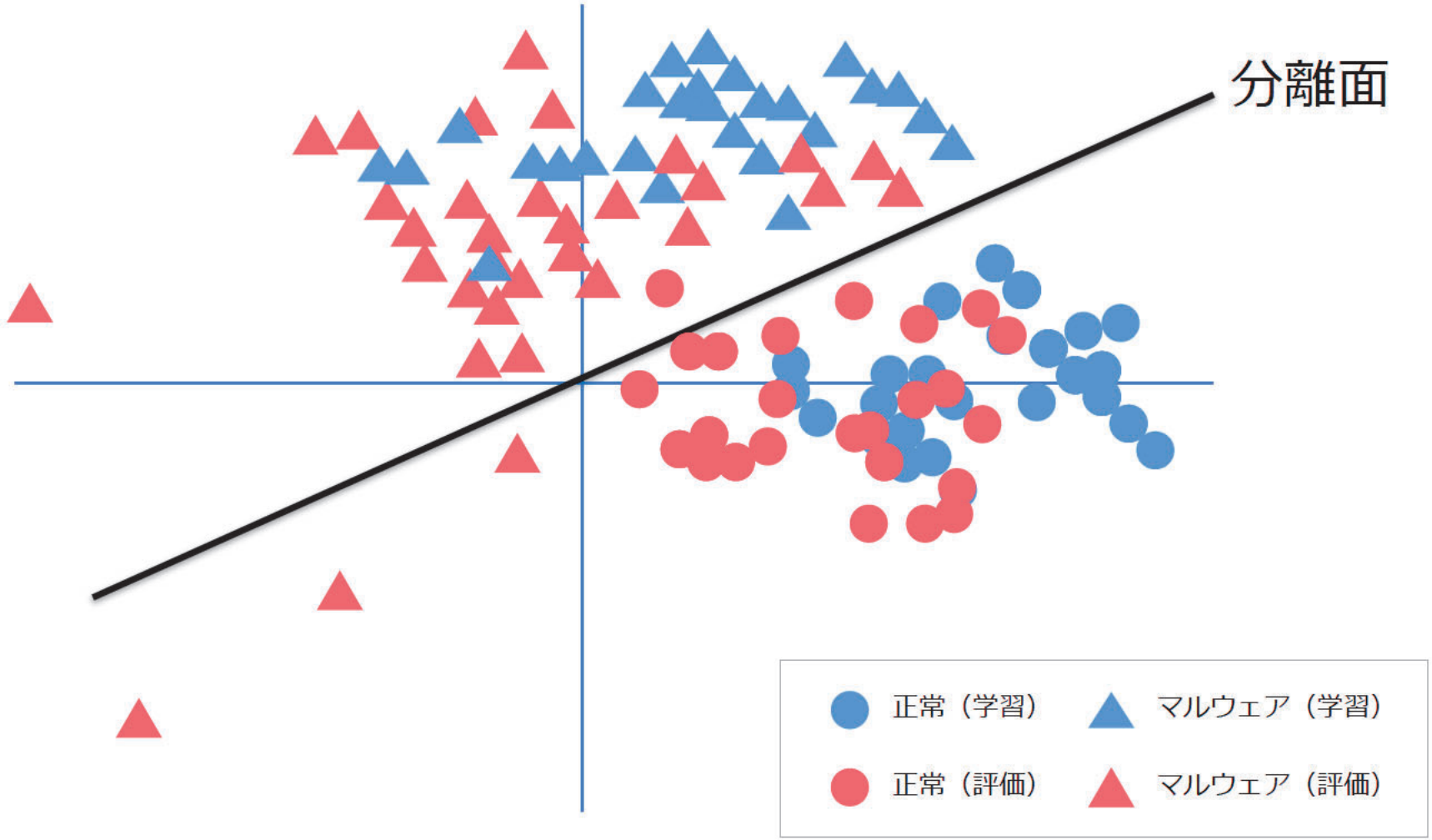


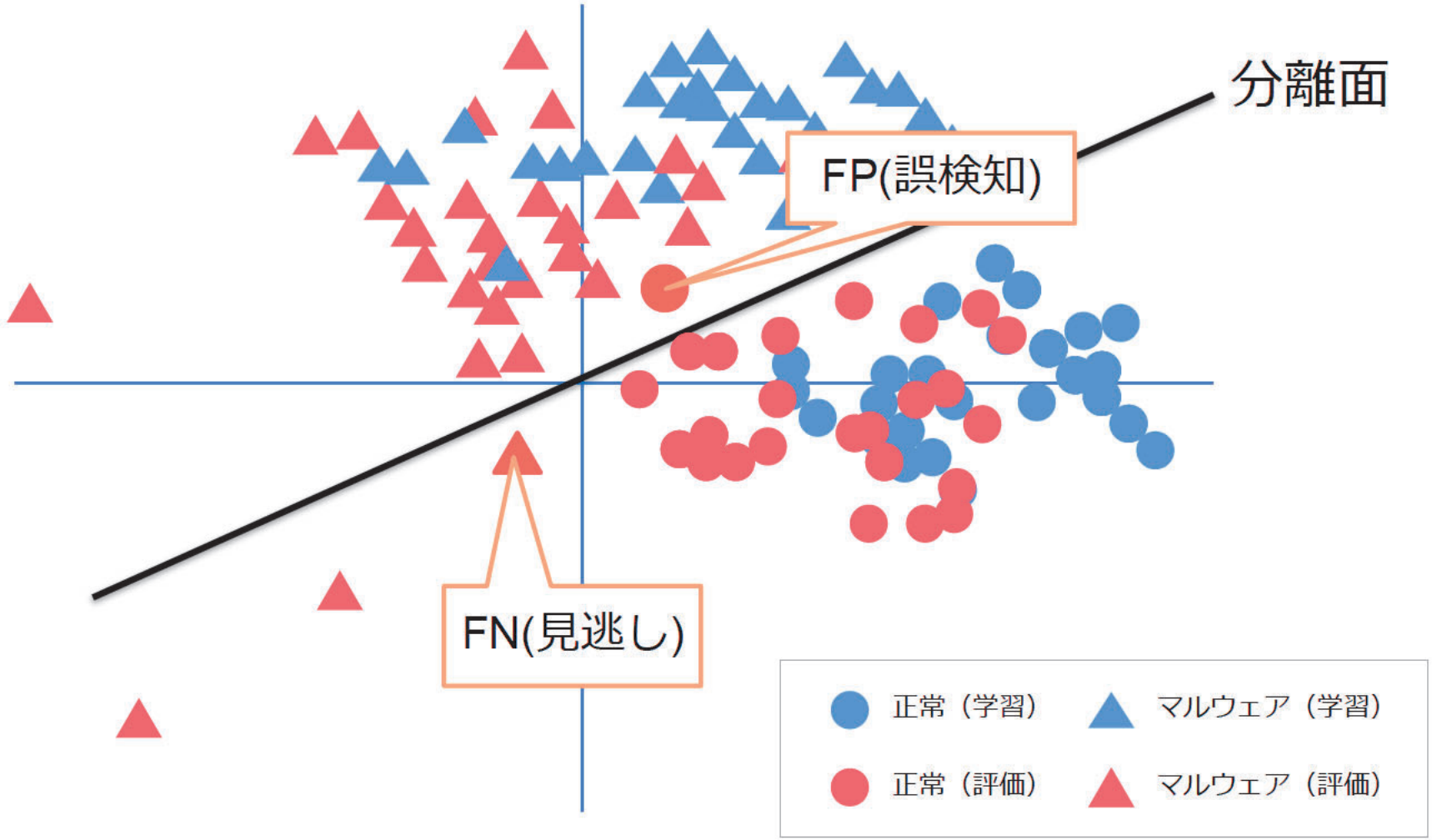
# 実験3(1/6) - 学習完了状態



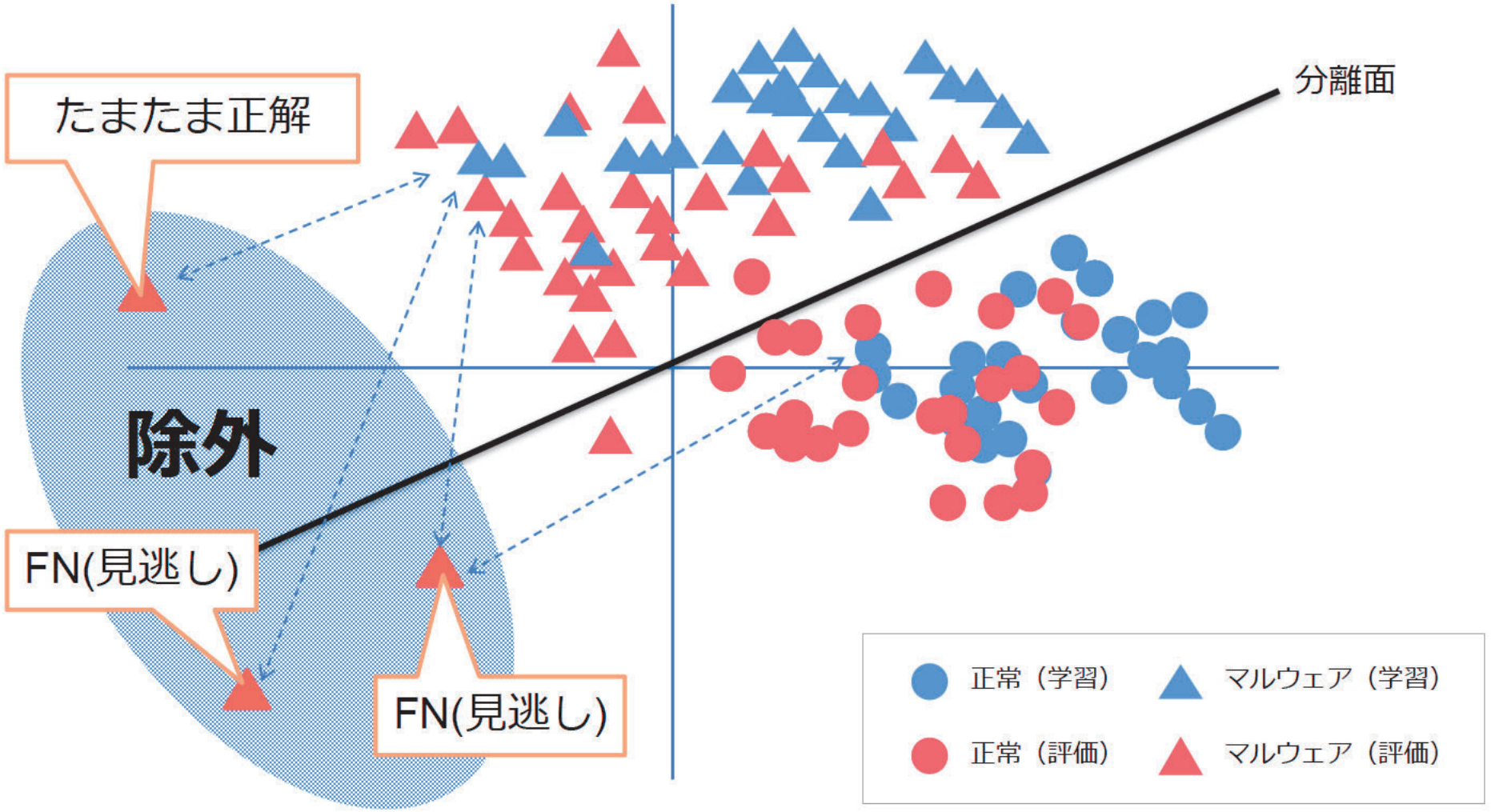
# 実験3(2/6) - 分類完了状態



# 実験3(2/6) - 分類完了状態

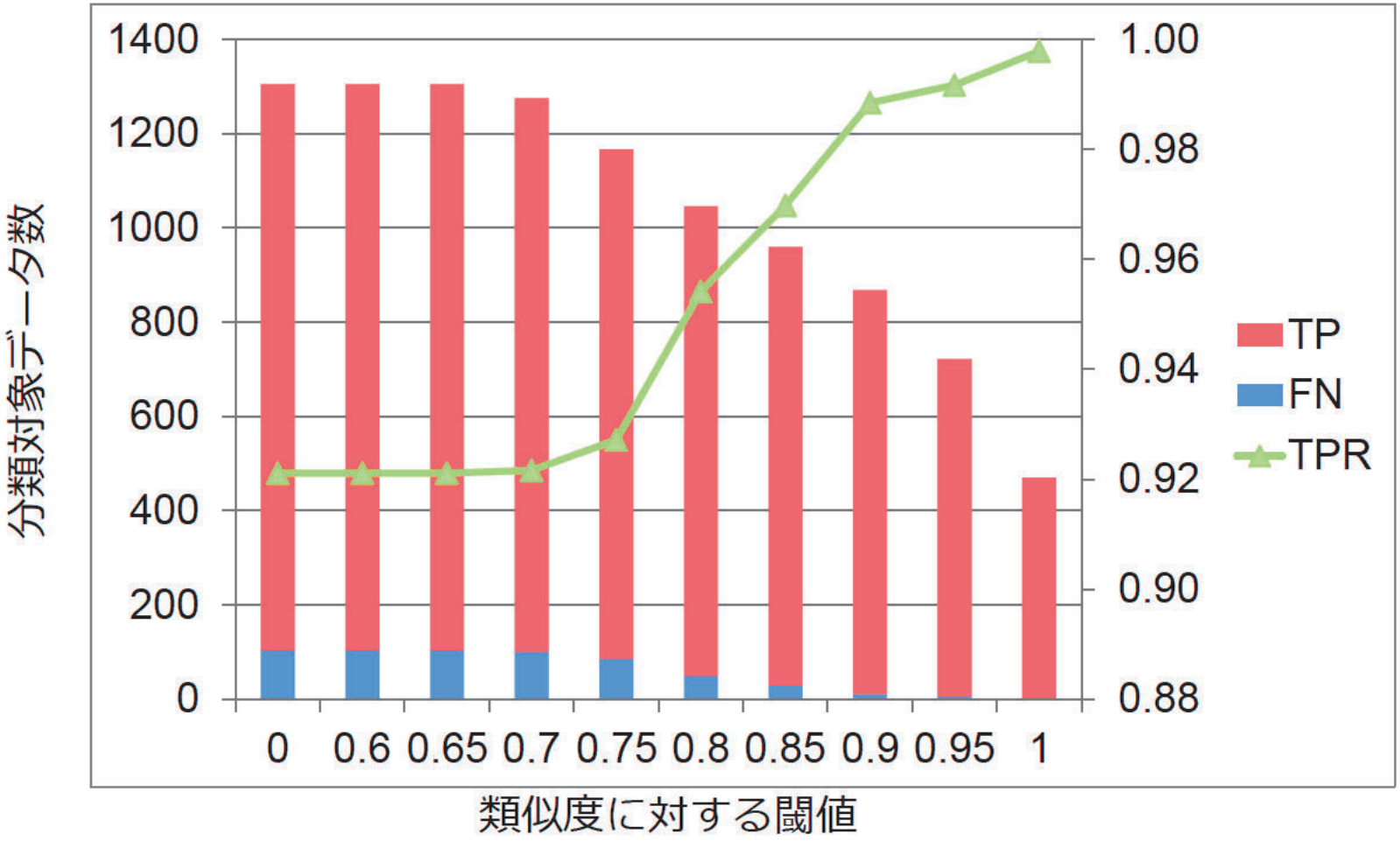


# 実験3(3/6) - 類似度の低い評価データ



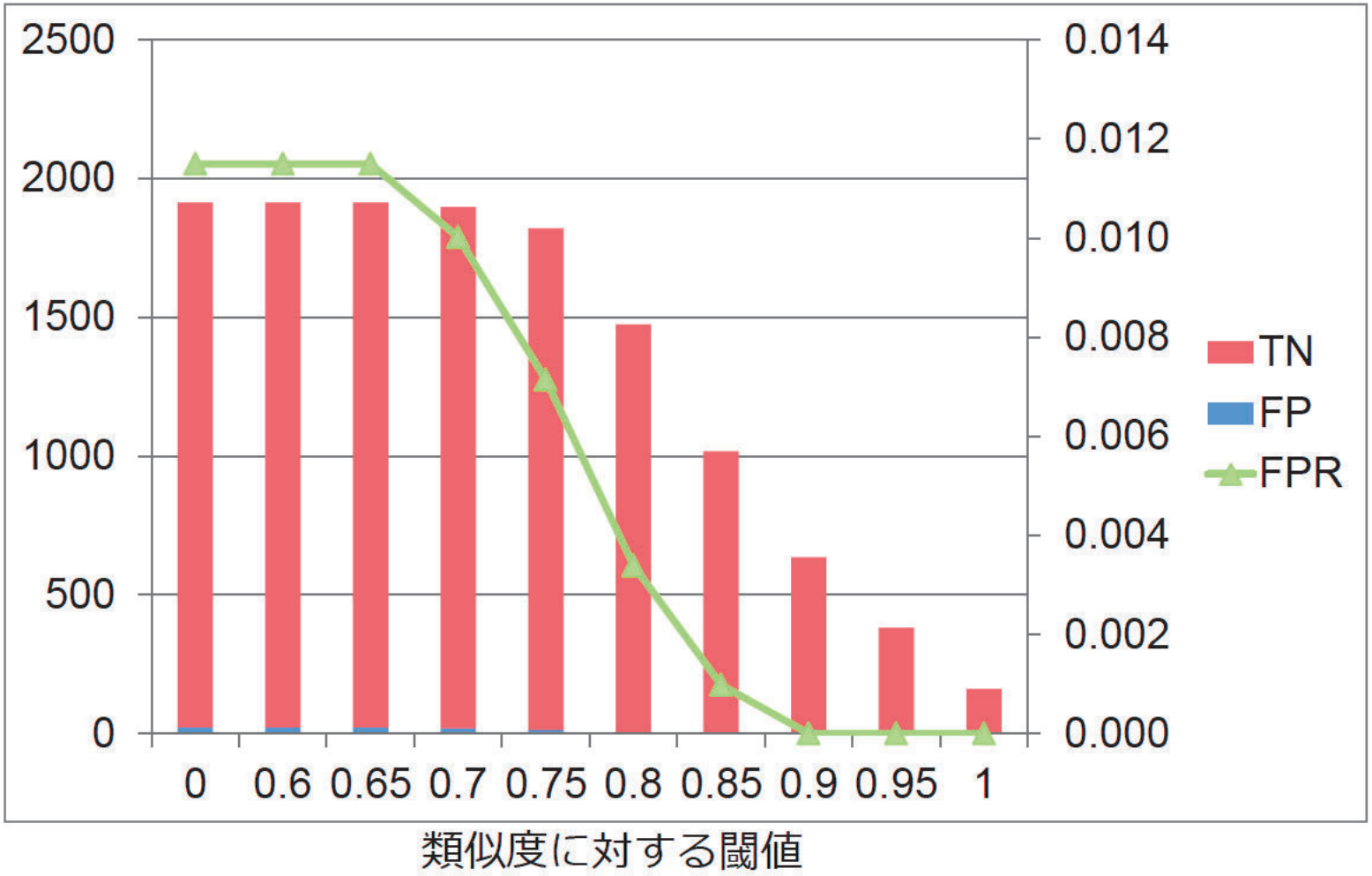
# 実験3(4/6) - TPRへの影響

## 分類対象の減少 ∝ TPR向上



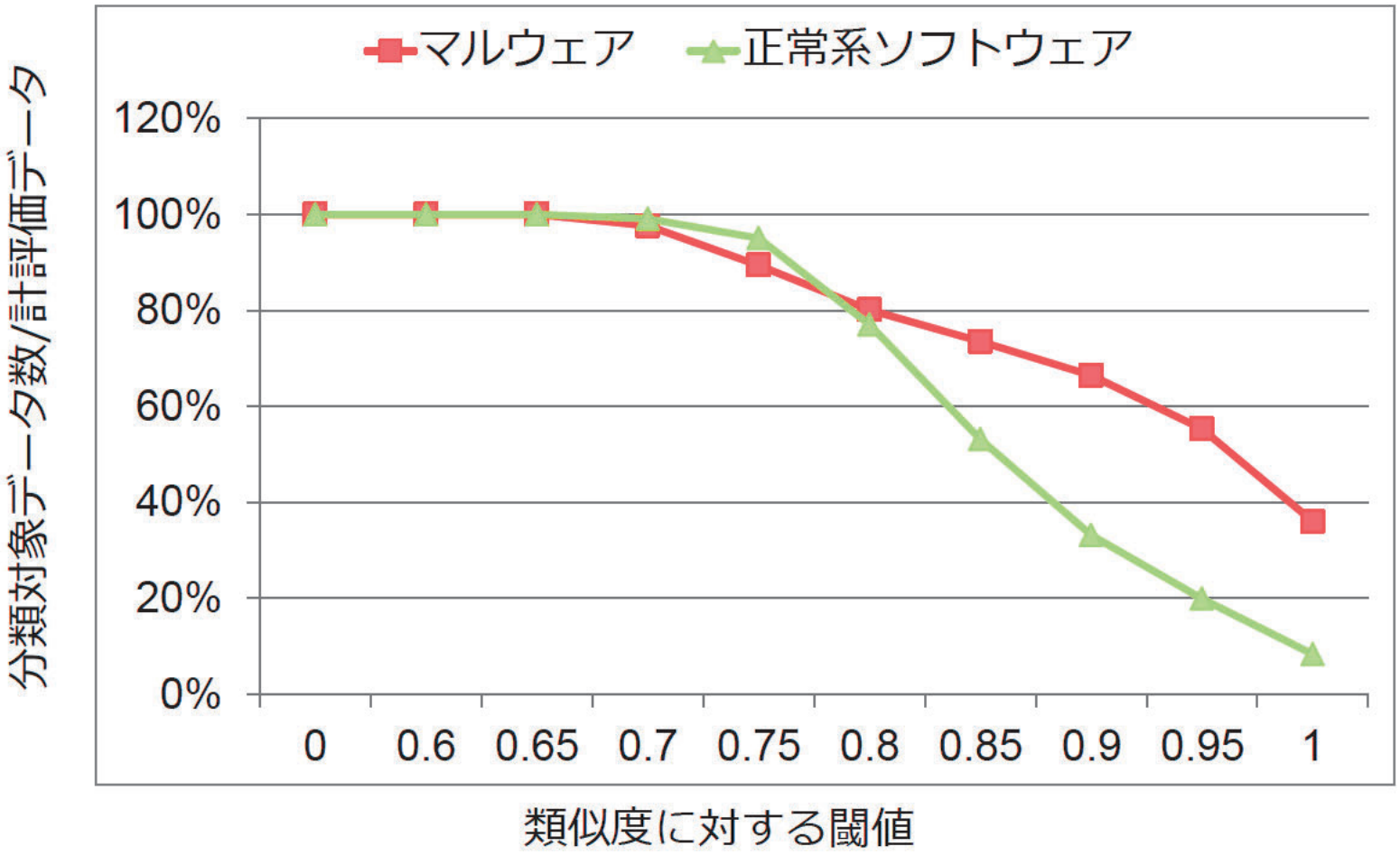
# 実験3(5/6) - FPRへの影響

## 分類対象の減少 ∝ FRP向上



# 実験3(6/6) - 分類対象データ数の推移

## 正常系の方がマルウェアに比べて減少率が高い



# 考察(1/3)

- 実際の適用シーン
  - マルウェアか正常系か分からないファイルを分類

- 実験３を適用した場合
  - 学習データ中に似たものがあれば分類対象に
  - 似たものがなければ対象外
    - 正解がマルウェアであればFalse Negative(見逃し)
    - 正解が正常系であればTrue Negative（結果的に正解）

- 上記より本質的には
  **「ユニークなマルウェアのTPR」**に関する問題
  （ユニークなマルウェアは見逃しがちに）

# 考察(2/3)

- 現状のようにマルウェアが多数の亜種を持つ場合、
  - 機械学習によるマルウェア検知は効果が期待できる
- 多数の亜種を持つ = 生成ツール
- 下記の調査が必要ではないか
  - マルウェア生成ツールの利用、普及動向
  - Anti-Machine Learning detection等の可能性

# 考察(3/3)

- 対象外としたマルウェアについて

1. 別の特徴を用いた分類を行う
2. データを増やす（ユニーク　→　非ユニークへの推移）
3. 機械学習以外の手法による検知を行う

# まとめ

- マルウェアと正常系では類似度の分布が異なる（実験1)
- これにより分類精度の悪化が発生する(実験2)
- 類似度の低いデータを除外した場合、
  ユニークなマルウェアのTPRが悪化する（実験3）
- 継続的なマルウェア、及び生成ツール等の動向調査が必要
- （正常系ソフトウェアを検知する技術が必要ではないか?)